

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:31:16 ; Search time 34 Seconds
(without alignments)
1478.747 Million cell updates/sec

Title: US-09-584-411c-22
Perfect score: 2017
Sequence: 1 MLHAANKGRKPSAEAGRPDP.....NDGDKEMVSNFTVVDGTI 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 497079 seqs, 131961718 residues

Total number of hits satisfying chosen parameters: 497079

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 2017 | 100.0 | 381 | 10 | US-09-808-602-6 |
| 2 | 2017 | 100.0 | 381 | 11 | US-09-977-418-22 |
| 3 | 2017 | 100.0 | 381 | 11 | US-09-977-033A-22 |
| 4 | 2017 | 100.0 | 381 | 11 | US-09-800-198-6 |
| 5 | 2017 | 100.0 | 381 | 12 | US-09-977-751C-22 |
| 6 | 1931 | 95.7 | 2724 | 10 | US-09-808-602-13 |
| 7 | 1931 | 95.7 | 2724 | 11 | US-09-800-198-13 |
| 8 | 1931 | 95.7 | 2723 | 10 | US-09-808-602-8 |
| 9 | 1931 | 95.7 | 2723 | 11 | US-09-800-198-8 |
| 10 | 1889.5 | 93.7 | 2764 | 10 | US-09-808-602-80 |
| 11 | 1889.5 | 93.7 | 2764 | 11 | US-09-800-198-68 |
| 12 | 1889.5 | 93.7 | 2765 | 10 | US-09-808-602-84 |
| 13 | 1889.5 | 93.7 | 2765 | 11 | US-09-800-198-72 |
| 14 | 1820.5 | 90.3 | 2802 | 10 | US-09-808-602-81 |
| 15 | 1820.5 | 90.3 | 2802 | 11 | US-09-800-198-69 |

| | | | | | | |
|----|-------|------|------|----|--------------------|-------------------|
| 16 | 793.5 | 39.3 | 2771 | 10 | US-09-808-602-82 | Sequence 82, Appl |
| 17 | 793.5 | 39.3 | 2771 | 11 | US-09-800-198-70 | Sequence 70, Appl |
| 18 | 790.5 | 39.2 | 560 | 15 | US-10-290-578-4 | Sequence 4, Appl1 |
| 19 | 790.5 | 39.2 | 768 | 9 | US-09-773-517-11 | Sequence 11, Appl |
| 20 | 790.5 | 39.2 | 768 | 9 | US-09-792-025-11 | Sequence 11, Appl |
| 21 | 790.5 | 39.2 | 768 | 9 | US-09-849-868-11 | Sequence 11, Appl |
| 22 | 790.5 | 39.2 | 768 | 10 | US-09-808-602-85 | Sequence 85, Appl |
| 23 | 790.5 | 39.2 | 768 | 15 | US-10-290-578-2 | Sequence 2, Appl1 |
| 24 | 658 | 32.6 | 501 | 15 | US-10-290-578-10 | Sequence 10, Appl |
| 25 | 129.5 | 6.4 | 731 | 14 | US-10-086-464-17 | Sequence 17, Appl |
| 26 | 124.5 | 6.2 | 731 | 14 | US-10-086-464-8 | Sequence 8, Appl1 |
| 27 | 124.5 | 6.2 | 4019 | 10 | US-09-738-973-425 | Sequence 425, App |
| 28 | 124.5 | 6.2 | 4019 | 10 | US-09-854-133-425 | Sequence 425, App |
| 29 | 124.5 | 6.2 | 4019 | 15 | US-10-144-649A-425 | Sequence 425, App |
| 30 | 122.5 | 6.1 | 279 | 14 | US-10-001-857-159 | Sequence 159, App |
| 31 | 121 | 6.0 | 5179 | 9 | US-09-922-217-1068 | Sequence 1068, Ap |
| 32 | 121 | 6.0 | 5179 | 10 | US-09-833-263-1068 | Sequence 1068, Ap |
| 33 | 121 | 6.0 | 5179 | 14 | US-10-025-380-1068 | Sequence 1068, Ap |
| 34 | 119 | 5.9 | 196 | 10 | US-09-989-920-224 | Sequence 224, App |
| 35 | 118.5 | 5.9 | 881 | 9 | US-09-816-860A-2 | Sequence 2, Appl1 |
| 36 | 117.5 | 5.8 | 884 | 15 | US-10-303-683-21 | Sequence 21, Appl |
| 37 | 117.5 | 5.8 | 894 | 11 | US-09-291-417-18 | Sequence 18, Appl |
| 38 | 117.5 | 5.8 | 894 | 15 | US-10-303-683-20 | Sequence 20, Appl |
| 39 | 117.5 | 5.8 | 1367 | 10 | US-09-801-368-108 | Sequence 108, App |
| 40 | 117 | 5.8 | 773 | 15 | US-10-225-630-6 | Sequence 6, Appl1 |
| 41 | 117 | 5.8 | 890 | 15 | US-10-158-684-4 | Sequence 4, Appl1 |
| 42 | 117 | 5.8 | 890 | 15 | US-10-158-711-4 | Sequence 4, Appl1 |
| 43 | 116 | 5.8 | 2971 | 15 | US-10-146-473-50 | Sequence 50, Appl |
| 44 | 115 | 5.7 | 495 | 15 | US-10-198-070-48 | Sequence 48, Appl |
| 45 | 115 | 5.7 | 833 | 15 | US-10-303-683-17 | Sequence 17, Appl |

ALIGNMENTS

RESULT 1

US-09-808-602-6
; Sequence 6, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NO. US20020155115A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-6

Query Match 100.0%; Score 2017; DB 10; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-150;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|----|---|-----|
| QY | 1 | MLHAANKGRKPSAEAGRPDPPTSSPSLLPSAOLPSSHNPVPSVCOMPLDSTSHQIMDT | 60 |
| DB | 1 | MLHAANKGRKPSAEAGRPDPPTSSPSLLPSAOLPSSHNPVPSVCOMPLDSTSHQIMDT | 60 |
| QY | 61 | NPDEEFSPNSYLLRACSGPQASSSGPPNNHHSQSTLRPLPPPHNHTLSHHSSANSLSNR | 120 |

Db 61 NPDEEFPNLYLRACSGPQASSSGPPNHHSSQSTLRPPPLPPPHNHTLSHHSSANSNLR 120
QY 121 NSLTNRSSQIHAPAPNDLATTPEVSQVLDQSWVNSNVPLETRHFLFKTSSTPLFSS 180
Db 121 NSLTNRSSQIHAPAPNDLATTPEVSQVLDQSWVNSNVPLETRHFLFKTSSTPLFSS 180
QY 181 SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
Db 181 SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
QY 241 YFTVPWSLKNSSIDSGEAEVGRVTOEVPFGVWRSQIHISQPOFLKFNISLGDALFGV 300
Db 241 YFTVPWSLKNSSIDSGEAEVGRVTOEVPFGVWRSQIHISQPOFLKFNISLGDALFGV 300
QY 301 YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQYLDVGLWHLAF 360
Db 301 YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQYLDVGLWHLAF 360
QY 361 YNDGKDKEMVSFNTVVDGTI 381
Db 361 YNDGKDKEMVSFNTVVDGTI 381

RESULT 2

US-09-977-418-22
; Sequence 22, Application US/09977418
; Publication No. US20030027158A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets et al
; TITLE OF INVENTION: No. US20030027158A1 polynucleotides and polypeptides encoded th
; FILE REFERENCE: 15966-552
; CURRENT APPLICATION NUMBER: US/09/977,418
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/584,411
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: USSN 60/189,810
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: USSN 60/191,158
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: USSN 60/193,086
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: USSN 60/201,388
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-977-418-22

Query Match 100.0%; Score 2017; DB 11; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-150;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHAANKGRKPSAEAGRPPTSPSLLPSAQLPSSHNPVSCOMPLDLSNTSHQIMDT 60
Db 1 MLHAANKGRKPSAEAGRPPTSPSLLPSAQLPSSHNPVSCOMPLDLSNTSHQIMDT 60
QY 61 NPDEEFPNLYLRACSGPQASSSGPPNHHSSQSTLRPPPLPPPHNHTLSHHSSANSNLR 120
Db 61 NPDEEFPNLYLRACSGPQASSSGPPNHHSSQSTLRPPPLPPPHNHTLSHHSSANSNLR 120
QY 121 NSLTNRSSQIHAPAPNDLATTPEVSQVLDQSWVNSNVPLETRHFLFKTSSTPLFSS 180
Db 121 NSLTNRSSQIHAPAPNDLATTPEVSQVLDQSWVNSNVPLETRHFLFKTSSTPLFSS 180
QY 181 SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
Db 181 SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
QY 241 YFTVPWSLKNSSIDSGEAEVGRVTOEVPFGVWRSQIHISQPOFLKFNISLGDALFGV 300

Db 241 YFTVPWSLKNSSIDSGEAEVGRVTOEVPFGVWRSQIHISQPOFLKFNISLGDALFGV 300
QY 301 YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQYLDVGLWHLAF 360
Db 301 YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQYLDVGLWHLAF 360
QY 361 YNDGKDKEMVSFNTVVDGTI 381
Db 361 YNDGKDKEMVSFNTVVDGTI 381

RESULT 3

US-09-977-033A-22
; Sequence 22, Application US/09977033A
; Publication No. US20030082554A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A
; APPLICANT: Fernandes, Elma
; APPLICANT: Herriman, John
; TITLE OF INVENTION: No. US20030082554A1 polynucleotides and human protein PRO228-like
; TITLE OF INVENTION: protein-like and human protein PRO228-like
; FILE REFERENCE: 15966-552 CON-S24
; CURRENT APPLICATION NUMBER: US/09/977,033A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 60/137,322
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/189,810
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/191,158
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/193,086
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/201,388
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 09/584,411
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-977-033A-22

Query Match 100.0%; Score 2017; DB 11; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-150;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHAANKGRKPSAEAGRPPTSPSLLPSAQLPSSHNPVSCOMPLDLSNTSHQIMDT 60
Db 1 MLHAANKGRKPSAEAGRPPTSPSLLPSAQLPSSHNPVSCOMPLDLSNTSHQIMDT 60
QY 61 NPDEEFPNLYLRACSGPQASSSGPPNHHSSQSTLRPPPLPPPHNHTLSHHSSANSNLR 120
Db 61 NPDEEFPNLYLRACSGPQASSSGPPNHHSSQSTLRPPPLPPPHNHTLSHHSSANSNLR 120
QY 121 NSLTNRSSQIHAPAPNDLATTPEVSQVLDQSWVNSNVPLETRHFLFKTSSTPLFSS 180
Db 121 NSLTNRSSQIHAPAPNDLATTPEVSQVLDQSWVNSNVPLETRHFLFKTSSTPLFSS 180
QY 181 SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
Db 181 SSPGYPLTSGVTYTPPPRLPRNTFSKAFKPKKPSKYCSWKCAALSAIAAALLAILLA 240
QY 241 YFTVPWSLKNSSIDSGEAEVGRVTOEVPFGVWRSQIHISQPOFLKFNISLGDALFGV 300
Db 241 YFTVPWSLKNSSIDSGEAEVGRVTOEVPFGVWRSQIHISQPOFLKFNISLGDALFGV 300
QY 301 YIRRGPPSHQAQYDFMERLDGKESVSPRRRSQITLVQNEAVFQYLDVGLWHLAF 360

Db 301 YIRRLPSPAQYDFMERLDGKESVSPRRRSIOTLVQNEAVFYQYLDVGLWHLAF 360
QY 361 YNDGDKEMVSPNTVVLDTGI 381
|||||
Db 361 YNDGDKEMVSPNTVVLDTGI 381

RESULT 4

US-09-800-198-6
; Sequence 6, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-198-6

Query Match 100.0%; Score 2017; DB 11; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-150;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHAANKGRKPSAEAGRIPTTSSPLLPSAQLPSSHNPVSCQMLLDSNTSHQIMDT 60
|||||
Db 1 MLHAANKGRKPSAEAGRIPTTSSPLLPSAQLPSSHNPVSCQMLLDSNTSHQIMDT 60
|||||
QY 61 NPDEEFSNSYLRLACSGPQQASSGPPNHHHSQSLRPLPPPHNHTLSHHSSANSNLR 120
|||||
Db 61 NPDEEFSNSYLRLACSGPQQASSGPPNHHHSQSLRPLPPPHNHTLSHHSSANSNLR 120
|||||
QY 121 NSLTNRRSQIHAPAPAPNDLATTPEVSQVQDSWVLSNVPLTRHFLKFTSSGSTPLFSS 180
|||||
Db 121 NSLTNRRSQIHAPAPAPNDLATTPEVSQVQDSWVLSNVPLTRHFLKFTSSGSTPLFSS 180
|||||
QY 181 SSPGYPLTSGTYVTPPRLPRNTFSRKAFKLKPKSKYCSWKCAALSATAAALLAILLA 240
|||||
Db 181 SSPGYPLTSGTYVTPPRLPRNTFSRKAFKLKPKSKYCSWKCAALSATAAALLAILLA 240
|||||
QY 241 YFIVPWSLKNSSIDSGEAEVGRVTVQEVPPGVFWRSQTHISQPFKFNISLGKDALFGV 300
|||||
Db 241 YFIVPWSLKNSSIDSGEAEVGRVTVQEVPPGVFWRSQTHISQPFKFNISLGKDALFGV 300
|||||
QY 301 YIRRLPSPAQYDFMERLDGKESVSPRRRSIOTLVQNEAVFYQYLDVGLWHLAF 360
|||||
Db 301 YIRRLPSPAQYDFMERLDGKESVSPRRRSIOTLVQNEAVFYQYLDVGLWHLAF 360
|||||
QY 361 YNDGDKEMVSPNTVVLDTGI 381
|||||
Db 361 YNDGDKEMVSPNTVVLDTGI 381

RESULT 5

US-09-977-751c-22
; Sequence 22, Application US/09977751C
; Publication No. US20030134430A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A

; APPLICANT: Fernandes, Elma
; APPLICANT: Herrman, John
; APPLICANT: Vernet, Corine
; TITLE OF INVENTION: No. US20030134430A1el Amino Acid Sequences for Human Caenorhab
; TITLE OF INVENTION: Polypeptides.
; FILE REFERENCE: 15966-552 CON S-40
; CURRENT APPLICATION NUMBER: US/09/977,751C
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 60/137,322
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/189,810
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/191,158
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/193,086
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/201,388
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 09/584,411
; PRIOR FILING DATE: 2000-05-31
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-977-751c-22

Query Match 100.0%; Score 2017; DB 12; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-150;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHAANKGRKPSAEAGRIPTTSSPLLPSAQLPSSHNPVSCQMLLDSNTSHQIMDT 60
|||||
Db 1 MLHAANKGRKPSAEAGRIPTTSSPLLPSAQLPSSHNPVSCQMLLDSNTSHQIMDT 60
|||||
QY 61 NPDEEFSNSYLRLACSGPQQASSGPPNHHHSQSLRPLPPPHNHTLSHHSSANSNLR 120
|||||
Db 61 NPDEEFSNSYLRLACSGPQQASSGPPNHHHSQSLRPLPPPHNHTLSHHSSANSNLR 120
|||||
QY 121 NSLTNRRSQIHAPAPAPNDLATTPEVSQVQDSWVLSNVPLTRHFLKFTSSGSTPLFSS 180
|||||
Db 121 NSLTNRRSQIHAPAPAPNDLATTPEVSQVQDSWVLSNVPLTRHFLKFTSSGSTPLFSS 180
|||||
QY 181 SSPGYPLTSGTYVTPPRLPRNTFSRKAFKLKPKSKYCSWKCAALSATAAALLAILLA 240
|||||
Db 181 SSPGYPLTSGTYVTPPRLPRNTFSRKAFKLKPKSKYCSWKCAALSATAAALLAILLA 240
|||||
QY 241 YFIVPWSLKNSSIDSGEAEVGRVTVQEVPPGVFWRSQTHISQPFKFNISLGKDALFGV 300
|||||
Db 241 YFIVPWSLKNSSIDSGEAEVGRVTVQEVPPGVFWRSQTHISQPFKFNISLGKDALFGV 300
|||||
QY 301 YIRRLPSPAQYDFMERLDGKESVSPRRRSIOTLVQNEAVFYQYLDVGLWHLAF 360
|||||
Db 301 YIRRLPSPAQYDFMERLDGKESVSPRRRSIOTLVQNEAVFYQYLDVGLWHLAF 360
|||||
QY 361 YNDGDKEMVSPNTVVLDTGI 381
|||||
Db 361 YNDGDKEMVSPNTVVLDTGI 381

RESULT 6

US-09-808-602-13
; Sequence 13, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S

QY 70 SYLLRACSGPQOASSGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSNLRSLNRRSQ 129
DB 222 SYLLRACSGPQOASSGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSNLRSLNRRSQ 281
QY 130 IHAPAPNDLATTESVOLQDSWVLSNPVLETRHFLFKTSSTGTPLEFSSSSPGYPLTS 189
DB 282 IHAPAPNDLATTESVOLQDSWVLSNPVLETRHFLFKTSSTGTPLEFSSSSPGYPLTS 341
QY 190 GTVYTPPRLPRNTFSRKAFLKKPKSKYCSWKCAALSAIAAALLLAILLAYFIVPWSLK 249
DB 342 GTVYTPPRLPRNTFSRKAFLKKPKSKYCSWKCAALSAIAAALLLAILLAYFIVPWSLK 401
QY 250 NSSIDSGAEVGRRTYQEVPGVFWRSQIHSQOFLKFNISLGKDALFGVYIRRLGPPS 309
DB 402 NSSIDSGAEVGRRTYQEVPGVFWRSQIHSQOFLKFNISLGKDALFGVYIRRLGPPS 461
QY 310 HAQYDFMERLDGKESVSVESPRRSTOTLVQNEAVFQYLDVGLHFLAFYNDGKDEM 369
DB 462 HAQYDFMERLDGKESVSVESPRRSTOTLVQNEAVFQYLDVGLHFLAFYNDGKDEM 521
QY 370 VSFNTVWLD 378
DB 522 VSFNTVWLD 530

RESULT 9

US-09-800-198-8
; Sequence 8, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Cornie AM
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-198-8

Query Match 95.7%; Score 1931; DB 11; Length 2733;
Best Local Similarity 98.9%; Pred. No. 7.4e-143;
Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 10 KPSAAGRPIPTSSPILLSAQLPSSHNPPVSCQMLLDSTSHQIMDTPNDEEFSN 69
DB 162 KSDDENGRPIPTSSPILLSAQLPSSHNPPVSCQMLLDSTSHQIMDTPNDEEFSN 221
QY 70 SYLLRACSGPQOASSGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSNLRSLNRRSQ 129
DB 222 SYLLRACSGPQOASSGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSNLRSLNRRSQ 281
QY 130 IHAPAPNDLATTESVOLQDSWVLSNPVLETRHFLFKTSSTGTPLEFSSSSPGYPLTS 189
DB 282 IHAPAPNDLATTESVOLQDSWVLSNPVLETRHFLFKTSSTGTPLEFSSSSPGYPLTS 341
QY 190 GTVYTPPRLPRNTFSRKAFLKKPKSKYCSWKCAALSAIAAALLLAILLAYFIVPWSLK 249
DB 342 GTVYTPPRLPRNTFSRKAFLKKPKSKYCSWKCAALSAIAAALLLAILLAYFIVPWSLK 401

QY 250 NSSIDSGAEVGRRTYQEVPGVFWRSQIHSQOFLKFNISLGKDALFGVYIRRLGPPS 309
DB 402 NSSIDSGAEVGRRTYQEVPGVFWRSQIHSQOFLKFNISLGKDALFGVYIRRLGPPS 461
QY 310 HAQYDFMERLDGKESVSVESPRRSTOTLVQNEAVFQYLDVGLHFLAFYNDGKDEM 369
DB 462 HAQYDFMERLDGKESVSVESPRRSTOTLVQNEAVFQYLDVGLHFLAFYNDGKDEM 521
QY 370 VSFNTVWLD 378
DB 522 VSFNTVWLD 530
RESULT 10
US-09-808-602-80
; Sequence 80, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 80
; LENGTH: 2764
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-808-602-80
Query Match 93.7%; Score 1889.5; DB 10; Length 2764;
Best Local Similarity 88.5%; Pred. No. 1.4e-139;
Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;
QY 10 KPSAAGRPIPTSSPILLSAQLPSSHNPPVSCQMLLDSTSHQIMDTPNDEEFSN 69
DB 162 KSDDENGRPIPTSSPILLSAQLPSSHNPPVSCQMLLDSTSHQIMDTPNDEEFSN 221
QY 70 SYLLRACSGPQOASSGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSNLRSLNRRSQ 129
DB 222 SYLLRACSGPQOASSGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSNLRSLNRRSQ 281
QY 130 IHAPAPNDLATTESVOLQDSWVLSNPVLETRHFLFKTSSTGTPLEFSSSSPGYPLTS 189
DB 282 IHAPAPNDLATTESVOLQDSWVLSNPVLETRHFLFKTSSTGTPLEFSSSSPGYPLTS 341
QY 190 GTVYTPPRLPRNTFSRKAFLKKPKSKYCSWKCAALSAIAAALLLAILLAYFI----- 243
DB 342 GTVYTPPRLPRNTFSRKAFLKKPKSKYCSWKCAALSAIAAALLLAILLAYFIAMHLG 401
QY 244 -----VPSLKNSSIDSGAEVGRRTYQEV 268
DB 402 LNWQLQPADGHTFNNGVTRGLPCNDVATVPSGKVPWSLKNSSIDSGAEVGRRTYQEV 461
QY 269 PGVFWRSQIHSQOFLKFNISLGKDALFGVYIRRLGPPSHAQYDFMERLDGKESV 328
DB 462 PGVFWRSQIHSQOFLKFNISLGKDALFGVYIRRLGPPSHAQYDFMERLDGKESV 521
QY 329 ESPRERRSTOTLVQNEAVFQYLDVGLHFLAFYNDGKEMVSFNTVWLD 378
DB 522 ESPRERRSTOTLVQNEAVFQYLDVGLHFLAFYNDGKEMVSFNTVWLD 571

RESULT 11
US-09-800-198-68
; Sequence 68, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Cornie AM
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 68
; LENGTH: 2764
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-800-198-68

Query Match 93.7%; Score 1889.5; DB 11; Length 2764;
Best Local Similarity 88.5%; Pred. No. 1.4e-139;
Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;
QY 10 KPSAAGRPPTPTSSSLPSAQLPSSHNPPVSCQPLDSTSHQIMDTNPDEEFSN 69
Db 162 KSDDDNGRPPTPTSSSLPSAQLPSSHNPPVSCQPLDSTSHQIMDTNPDEEFSN 221
QY 70 SYLLRACSGPQQAASSGGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSLSLNRSQ 139
Db 222 SYLLRACSGPQQAASSGGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSLSLNRSQ 281
QY 130 IHAPAPNDLATTPEVSQVQLDSWVLSNVPETRLPFLKTSSTGSPPLFSSSSPGYPLTS 189
Db 282 IHAPAPNDLATTPEVSQVQLDSWVLSNVPETRLPFLKTSSTGSPPLFSSSSPGYPLTS 341
QY 190 GTVYTPPRLPNTFSRKAFKLPKSKYCSWKAALSAIAAALLAILLAYFI----- 243
Db 342 GTVYTPPRLPNTFSRKAFKLPKSKYCSWKAALSAIAAALLAILLAYFIAMHLG 401
QY 244 -----VPWSLKNSSIDSGEAEVGRVYQEV 268
Db 402 LNWQLPADGHTFNNGVRTGLPGNDVAVTPSGGKVPWSLKNSSIDSGEAEVGRVYQEV 461
QY 269 PGVFWRSQIHSQPFQFLAFNISGLKDALFGVYIRGLPPSHQAQYDFMERLDGKEKWSV 328
Db 462 PGVFWRSQIHSQPFQFLAFNISGLKDALFGVYIRGLPPSHQAQYDFMERLDGKEKWSV 521
QY 329 ESPRRRSIQTLVQNEAVFYQYLDVGLWHLAFYNDGKDKEMVSFNTVWLD 378
Db 522 ESPRRRSIQTLVQNEAVFYQYLDVGLWHLAFYNDGKDKEMVSFNTVWLD 571

RESULT 12
US-09-808-602-84
; Sequence 84, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Cornie A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu

; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NO. US20020155115A1el Proteins and Nuclec Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 84
; LENGTH: 2765
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-808-602-84

Query Match 93.7%; Score 1889.5; DB 10; Length 2765;
Best Local Similarity 88.5%; Pred. No. 1.4e-139;
Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;
QY 10 KPSAAGRPPTPTSSSLPSAQLPSSHNPPVSCQPLDSTSHQIMDTNPDEEFSN 69
Db 162 KSDDDNGRPPTPTSSSLPSAQLPSSHNPPVSCQPLDSTSHQIMDTNPDEEFSN 221
QY 70 SYLLRACSGPQQAASSGGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSLSLNRSQ 129
Db 222 SYLLRACSGPQQAASSGGPPNHHQSOTLRPPLPPPHNHTLSHHSSANSLSLNRSQ 281
QY 130 IHAPAPNDLATTPEVSQVQLDSWVLSNVPETRLPFLKTSSTGSPPLFSSSSPGYPLTS 189
Db 282 IHAPAPNDLATTPEVSQVQLDSWVLSNVPETRLPFLKTSSTGSPPLFSSSSPGYPLTS 341
QY 190 GTVYTPPRLPNTFSRKAFKLPKSKYCSWKAALSAIAAALLAILLAYFI----- 243
Db 342 GTVYTPPRLPNTFSRKAFKLPKSKYCSWKAALSAIAAALLAILLAYFIAMHLG 401
QY 244 -----VPWSLKNSSIDSGEAEVGRVYQEV 268
Db 402 LNWQLPADGHTFNNGVRTGLPGNDVAVTPSGGKVPWSLKNSSIDSGEAEVGRVYQEV 461
QY 269 PGVFWRSQIHSQPFQFLAFNISGLKDALFGVYIRGLPPSHQAQYDFMERLDGKEKWSV 328
Db 462 PGVFWRSQIHSQPFQFLAFNISGLKDALFGVYIRGLPPSHQAQYDFMERLDGKEKWSV 521
QY 329 ESPRRRSIQTLVQNEAVFYQYLDVGLWHLAFYNDGKDKEMVSFNTVWLD 378
Db 522 ESPRRRSIQTLVQNEAVFYQYLDVGLWHLAFYNDGKDKEMVSFNTVWLD 571

RESULT 13
US-09-800-198-72
; Sequence 72, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Cornie AM
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72

```
; LENGTH: 2765
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-800-198-72

Query Match
Best Local Similarity 93.7%; Score 1889.5; DB 11; Length 2765;
Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;

QY 10 KPSAAGRIPTSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 69
DB 162 KSDDDNGRPIPTSSSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 221
QY 70 SYLLRACSGPOAASSGPPNHSQSTLRPPPPHNTLSHHSSANSLNRSNTNRSQ 129
DB 222 SYLLRACSGPOAASSGPPNHSQSTLRPPPPHNTLSHHSSANSLNRSNTNRSQ 281
QY 130 IHAPAPNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 189
DB 282 IHAPAPNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 341
QY 190 GTVYTPPPRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 243
DB 342 GTVYTPPPRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 401
QY 244 -----VPSLNKSSIDSGEAEVGRRTQEV 268
DB 402 LNWQLQPADGHTFNGVRTGLPGNDVATVPSGKVPWPSLNKSSIDSGEAEVGRRTQEV 461
QY 269 PPGVFWRSQIHSIQPOFLKFNISLQKDALFGVYIRRGLPSSHAQYDFMERLDGKESV 328
DB 462 PPGVFWRSQIHSIQPOFLKFNISLQKDALFGVYIRRGLPSSHAQYDFMERLDGKESV 521
QY 329 ESPRRRSIOTLVQNEAVFOYLDVGLWHLAFYNDGDKEMVSNFTVILD 378
DB 522 ESPRRRSIOTLVQNEAVFOYLDVGLWHLAFYNDGDKEMVSNFTVILD 571

RESULT 14
US-09-808-602-81
; Sequence 81, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NO. US20020155115A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 2802
; TYPE: PRT
; ORGANISM: Gallus gallus
US-09-808-602-81

Query Match
Best Local Similarity 90.3%; Score 1820.5; DB 10; Length 2802;
Matches 346; Conservative 9; Mismatches 7; Indels 41; Gaps 1;

QY 17 RPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 76
DB 197 RPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 256
QY 77 SGPOQASSGPPNHSQSTLRPPPPHNTLSHHSSANSLNRSNTNRSQIHAPAPA 136
DB 257 SGPOQASSGPPNHSQSTLRPPPPHNTLSHHSSANSLNRSNTNRSQIHAPAPA 316
QY 137 PNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 196
DB 317 PNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 376
QY 197 PRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 244
DB 377 PRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 436
QY 245 -----PMSLNKSSIDSGEAEVGRRTQEV 275
DB 437 ADGHTFNSGLRPGAAGAEAGAAAPGAGRPVWTRNSIDSGETEVRKVTQEVPGVFW 496
QY 276 SQIHSIQPOFLKFNISLQKDALFGVYIRRGLPSSHAQYDFMERLDGKESV 335
DB 497 SQIHSIQPOFLKFNISLQKDALFGVYIRRGLPSSHAQYDFMERLDGKESV 556
QY 336 SIQTLVQNEAVFOYLDVGLWHLAFYNDGDKEMVSNFTVILD 378
DB 557 SIQTLVQNEAVFOYLDVGLWHLAFYNDGDKEMVSNFTVILD 599

RESULT 15
US-09-800-198-69
; Sequence 69, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
; LENGTH: 2802
; TYPE: PRT
; ORGANISM: Gallus gallus
US-09-800-198-69

Query Match
Best Local Similarity 90.3%; Score 1820.5; DB 11; Length 2802;
Matches 346; Conservative 9; Mismatches 7; Indels 41; Gaps 1;

QY 17 RPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 76
DB 197 RPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDNTSHQIMDTPDEEFSN 256
QY 77 SGPOQASSGPPNHSQSTLRPPPPHNTLSHHSSANSLNRSNTNRSQIHAPAPA 136
DB 257 SGPOQASSGPPNHSQSTLRPPPPHNTLSHHSSANSLNRSNTNRSQIHAPAPA 316
QY 137 PNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 196
DB 317 PNDLATTPEVSQVQDSWVLSNVPLETRHFLFKTSSTGTPLFSSSSPGYPLTS 376
QY 197 PRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 244
DB 377 PRLPRNTFSRKAFKPKSKYCSWKCAALSAIAAALLAILLAYFIV 436
```

```
QY 245 -----PWSLKNSIDSGEAEVGRRTQEVPPGVFWR 275
Db 437 ADGHTFNGLRPGAACAEDEGAAAPPAGRGFWTRNRSIDSGETEVGRKVTQEVPPGVFWR 496
QY 276 SQIHISQOFLKFENISLGDALFQVYIRGLPPSHAQYDFMERLDGKWKSWVVEPRRR 335
Db 497 SQIHISQOFLKFENISLGDALFQVYIRGLPPSHAQYDFMERLDGKWKSWVVEPRRR 556
QY 336 SIQTLVQNEAVFQYLDVGLHFLAFYNDGKDKEMVSFNTVWLD 378
Db 557 SIQTLVQNEAVFQYLDVGLHFLAFYNDGKDKREVSFSTVILD 599
```

Search completed: August 26, 2003, 14:33:31
Job time : 35 secs

| Result No. | Score | Query | | Length | DB | ID | Description |
|------------|-------|-------|------|--------|----------|--------------------|-------------|
| | | Match | | | | | |
| 1 | 2017 | 100.0 | 381 | 22 | AAU08679 | Human FCTR3a polyp | |
| 2 | 2017 | 100.0 | 381 | 22 | AAB61141 | Human NOV11 protei | |
| 3 | 2005 | 99.4 | 429 | 22 | AAU00392 | Human secreted pro | |
| 4 | 1931 | 95.7 | 2724 | 22 | AAU08661 | Human FCTR3f polyp | |
| 5 | 1931 | 95.7 | 2733 | 22 | AAU08680 | Human FCTR3b polyp | |
| 6 | 992.5 | 49.2 | 865 | 22 | ABG03234 | Novel human diagno | |
| 7 | 976.5 | 48.4 | 2721 | 23 | ABP53567 | Human NOV15b prote | |
| 8 | 972.5 | 48.2 | 2613 | 23 | ABP53589 | Human NOV15d prote | |
| 9 | 972.5 | 48.2 | 2628 | 23 | ABP53588 | Human NOV15c prote | |

SUMMARIES

XX

PD 13-SEP-2001.
 XX 05-MAR-2001; 2001WO-US07160.
 XX 03-MAR-2000; 2000US-186592P.
 PR 03-MAR-2000; 2000US-186718P.
 PR 06-MAR-2000; 2000US-187293P.
 PR 06-MAR-2000; 2000US-187294P.
 PR 17-MAR-2000; 2000US-190400P.
 PR 07-APR-2000; 2000US-196018P.
 PR 03-JAN-2001; 2001US-259548P.
 XX (CURA-) CURAGEN CORP.
 XX Vernet CAM, Fernandes E, Shinkets RA, Herrmann JL, Majumder K;
 PI Macdougall J, Mishra V, Mezes PS, Rastelli L;
 XX WPI: 2001-596837/67.
 DR N-PSDB; AAS14084.
 XX Novel polypeptides designated as FCTR polypeptides, useful in
 PT detection, prevention and treatment of a broad range of pathological
 PT states
 XX
 PS Claim 1; Page 33; 215pp; English.
 XX The invention relates to human FCTR polypeptides, FCTR1-FCTR7, and the
 CC nucleic acids encoding them. These sequences are useful for the treatment
 CC or prevention of numerous disorders including myelogenous leukaemia,
 CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal
 CC allolimmune thrombocytopenia, neurological disorders, neurodegenerative
 CC disorders, nerve trauma, familial myelodysplastic syndrome,
 CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial
 CC myelodysplastic syndrome, mental health conditions, immunological
 CC disorders, allergy and infection, bronchial asthma, Avellino type
 CC eosinophilia, lung diseases, reproductive disorders, infertility, male
 CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,
 CC desmoid disease, turcot syndrome, liver cirrhosis, hepatitis C, gastric
 CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni
 CC infection Spino-cerebellar ataxia, Plasmodium falciparum parasitaemia,
 CC Corneal dystrophy-Greenow type I, Corneal dystrophy-lattice type I and
 CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR3a, a
 CC neurestin-like protein.
 XX
 SQ Sequence 381 AA;
 Query Match 100.0%; Score 2017; DB 22; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.1e-159;
 Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLHAANKGRKPSAEAGRIPTTSSPLPSAQLPSSHPNPVSCQPLDLSNTSHQIMDT 60
 DB 1 MLHAANKGRKPSAEAGRIPTTSSPLPSAQLPSSHPNPVSCQPLDLSNTSHQIMDT 60
 QY 61 NPDEFFSNVLLRACSGPQQAASSGPPNHHQSQTLPRLPPPHNHTLSHHSSANSLNR 120
 DB 61 NPDEFFSNVLLRACSGPQQAASSGPPNHHQSQTLPRLPPPHNHTLSHHSSANSLNR 120
 QY 121 NSLTNRSSQIHAPAPAPNDLATTPEVOLQDSWVLSNVPLETRFLFKTSSTGSPPLFSS 180
 DB 121 NSLTNRSSQIHAPAPAPNDLATTPEVOLQDSWVLSNVPLETRFLFKTSSTGSPPLFSS 180
 QY 181 SSPGYPLSGVYTPPPRLPNTTSKAFKLPKSKYCSWKAALSAIAAALLAILLA 240
 DB 181 SSPGYPLSGVYTPPPRLPNTTSKAFKLPKSKYCSWKAALSAIAAALLAILLA 240
 QY 241 YFIVPWSLKNSSIDSGEAEVGRVTOEYPPGVFWRSQIHISQOFLKFNISLGKDALGV 300
 DB 241 YFIVPWSLKNSSIDSGEAEVGRVTOEYPPGVFWRSQIHISQOFLKFNISLGKDALGV 300
 QY 301 YIRGLPSPHAQYDFMERLDGKESVSPRRRSIQTLVQNEAVFYQYLDVGLWHLAF 360
 DB 301 YIRGLPSPHAQYDFMERLDGKESVSPRRRSIQTLVQNEAVFYQYLDVGLWHLAF 360

QY 361 YNDGKDKEMVSFNTVVLDTGTI 381
 DB 361 YNDGKDKEMVSFNTVVLDTGTI 381
 RESULT 2
 AAB61141
 ID AAB61141 standard; Protein; 381 AA.
 XX
 AC AAB61141;
 XX
 DT 30-MAR-2001 (first entry)
 XX Human NOV11 protein.
 DE
 XX Human; NOVX: antiinflammatory; cytostatic; neuroprotective;
 KW cerebroprotective; immunomodulator; vulnarary; vasotropic; gene therapy;
 KW hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
 KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200075321-A2.
 XX
 PD 14-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US15303.
 XX
 PR 03-JUN-1999; 99US-0137322.
 PR 16-MAR-2000; 2000US-0189810.
 PR 22-MAR-2000; 2000US-0191158.
 PR 30-MAR-2000; 2000US-0193086.
 PR 31-MAY-2000; 2000US-0137322.
 XX
 XX (CURA-) CURAGEN CORP.
 XX Shinkets RA, Fernandes E, Herrman J, Vernet C;
 XX WPI: 2001-102403/11.
 DR N-PSDB; AAF27859.
 XX
 PT New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug
 PT target for treating immune, proliferative and metabolic diseases and
 PT wound healing
 XX
 PS Claim 1; Page 42-44; 194pp; English.
 XX
 CC The present sequence is a new isolated polypeptide (NOVX). The NOVX
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
 CC treating or preventing NOVX-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.
 CC These NOVX-associated disorders include hyperplasias, tumours,
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVX is useful as
 CC a diagnostic marker or prognostic marker, protein therapeutic and
 CC antibody target or small molecule drug target to treat disorders in the
 CC immune response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns).
 XX
 SQ Sequence 381 AA;
 Query Match 100.0%; Score 2017; DB 22; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.1e-159;
 Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLHAANKGRKPSAEAGRIPTTSSPLPSAQLPSSHPNPVSCQPLDLSNTSHQIMDT 60

|||||
Db 1 MLHANKRKPSAEAGRIPTTSSPSLLPSAQLPSSHNPPVPVSCOMPLDSTSHQIMDT 60
QY 61 NPDEEFPNSYLLRACSGPQOASSGPPNHHSSQSTLRPPLPPPHNHTLSHHSSANSNLR 120
Db 61 NPDEEFPNSYLLRACSGPQOASSGPPNHHSSQSTLRPPLPPPHNHTLSHHSSANSNLR 120
QY 121 NSLNRSSQIHAAPAPANDLATTPEVSQLOQSWLNSVPLETRHFLFKTSSTGTPFSS 180
Db 121 NSLNRSSQIHAAPAPANDLATTPEVSQLOQSWLNSVPLETRHFLFKTSSTGTPFSS 180
QY 181 SSPGYPLTSGTVYTPPPRLPNTFSRKAFLKPKSKYCSWKCAALSAAIAALLAILLA 240
Db 181 SSPGYPLTSGTVYTPPPRLPNTFSRKAFLKPKSKYCSWKCAALSAAIAALLAILLA 240
QY 241 YFIVPWSLKNSSIDSGEAEVGRRTVQEPVPGVFWRSQIHISQOFLKFNISLKGKDALFGV 300
Db 241 YFIVPWSLKNSSIDSGEAEVGRRTVQEPVPGVFWRSQIHISQOFLKFNISLKGKDALFGV 300
QY 301 YIRGLPPSHAQYDFMERLDGKEKWSVVEPRRRIOTLVQNEAVFYQYLDVGLWHLAF 360
Db 301 YIRGLPPSHAQYDFMERLDGKEKWSVVEPRRRIOTLVQNEAVFYQYLDVGLWHLAF 360
QY 361 YNDGDKKEMVSFNTVLDGTI 381
Db 361 YNDGDKKEMVSFNTVLDGTI 381

RESULT 3

AAU00392
ID AAU00392 standard; Protein; 429 AA.

XX AC AAU00392;

XX DT 04-JUL-2001 (first entry)

XX DE Human secreted protein, POLY4.

XX KW Human secreted protein; therapeutic; diagnostic; human; cancer.

XX OS Homo sapiens.

XX PN WO200119856-A2.

XX PD 22-MAR-2001.

XX PF 13-SEP-2000; 2000WO-US25106.

XX PR 13-SEP-1999; 99US-0153629.

XX PR 16-SEP-1999; 99US-0154520.

XX PR 20-SEP-1999; 99US-0154762.

XX PR 13-OCT-1999; 99US-0159231.

XX PR 12-SEP-2000; 2000US-0659634.

XX PA (CURA-) CURAGEN CORP.

XX PI Shinkets RA, Fernandes E, Herrmann JL, Liu X, Yang M, Boldog FL;

XX DR WPI; 2001-244781/25.

XX DR N-PSDB; AAS01213.

XX XX
PT New POLYX polypeptide useful for treating or preventing a POLYX
associated disorder, e.g. cancer -
PS Claim 9; Page 11-13; 152pp; English.
XX
CC The sequence represents the amino acid sequence of human secreted
protein, POLY4. POLYX nucleic acids, polypeptides and antibodies to POLYX
can be used for treating or preventing a POLYX associated disorder in a
subject, preferably a human. These can be used in the manufacture of a
medicament for treating a syndrome associated with a human disease
selected from a POLYX-associated disorder, where the therapeutic is a
POLYX polypeptide, a POLYX nucleotide or a POLYX antibody. They may also

CC be used to screen for a modulator of activity, or latency, or
CC predisposition to a POLYX associated disorder, e.g. cancer.

XX SQ Sequence 429 AA;

Query Match 99.4%; Score 2005; DB 22; Length 429;
Best Local Similarity 99.2%; Pred. No. 1.3e-158;
Matches 378; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLHANKRKPSAEAGRIPTTSSPSLLPSAQLPSSHNPPVPVSCOMPLDSTSHQIMDT 60
Db 1 MLHANKRKPSAEAGRIPTTSSPSLLPSAQLPSSHNPPVPVSCOMPLDSTSHQIMDT 60

QY 61 NPDEEFPNSYLLRACSGPQOASSGPPNHHSSQSTLRPPLPPPHNHTLSHHSSANSNLR 120
Db 61 NPDEEFPNSYLLRACSGPQOASSGPPNHHSSQSTLRPPLPPPHNHTLSHHSSANSNLR 120

QY 121 NSLNRSSQIHAAPAPANDLATTPEVSQLOQSWLNSVPLETRHFLFKTSSTGTPFSS 180
Db 121 NSLNRSSQIHAAPAPANDLATTPEVSQLOQSWLNSVPLETRHFLFKTSSTGTPFSS 180

QY 181 SSPGYPLTSGTVYTPPPRLPNTFSRKAFLKPKSKYCSWKCAALSAAIAALLAILLA 240
Db 181 SSPGYPLTSGTVYTPPPRLPNTFSRKAFLKPKSKYCSWKCAALSAAIAALLAILLA 240

QY 241 YFIVPWSLKNSSIDSGEAEVGRRTVQEPVPGVFWRSQIHISQOFLKFNISLKGKDALFGV 300
Db 241 YFIVPWSLKNSSIDSGEAEVGRRTVQEPVPGVFWRSQIHISQOFLKFNISLKGKDALFGV 300

QY 301 YIRGLPPSHAQYDFMERLDGKEKWSVVEPRRRIOTLVQNEAVFYQYLDVGLWHLAF 360
Db 301 YIRGLPPSHAQYDFMERLDGKEKWSVVEPRRRIOTLVQNEAVFYQYLDVGLWHLAF 360

QY 361 YNDGDKKEMVSFNTVLDGTI 381
Db 361 YNDGDKKEMVSFNTVLDGVS 381

RESULT 4

AAU08681
ID AAU08681 standard; Protein; 2724 AA.

XX AC AAU08681;

XX DT 18-DEC-2001 (first entry)

XX DE Human FCTR3f polypeptide sequence.

XX KW Human; FCTR3f; myelogenous leukaemia; carcinoma; melanoma; glioma;
astrocytoma; congenital neonatal alloimmune thrombocytopenia; infection;
neurological disorder; neurodegenerative disorders; nerve trauma;
familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;
demyelinating Gardner syndrome; familial myelodysplastic syndrome;
mental health condition; immunological disorder; allergy; infertility;
bronchial asthma; Aveilino type eosinophilia; lung disease; deafness;
reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;
desmoid disease; turcot syndrome; liver cirrhosis; hepatitis C; virucide;
gastric disorders; pancreatic disease; Schistosoma mansoni infection;
Spino cerebellar ataxia; Plasmodium falciparum parasitaemia; diabetes;
Corneal dystrophy-Greenow type I; Corneal dystrophy-lattice type I;
Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive;
antidiabetic; antidiabetic; antidiabetic; antidiabetic;
gynaecological; antidiabetic; hepatotropic; virucide; ophthalmological;
gene therapy; FCTR3a; neurestin-like protein; FCTR3f.

XX OS Homo sapiens.

XX PN WO200166747-A2.

XX PD 13-SEP-2001.

XX PF 05-MAR-2001; 2001WO-US07160.

CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni
 CC infection Spinocerebellar ataxia, Plasmodium falciparum parasitaemia,
 CC Corneal dystrophy-Greenow type I, Corneal dystrophy-lattice type I and
 CC Rels-Bucklers corneal dystrophy. This sequence represents FCtr3b, a
 CC neurestin-like protein.
 XX
 SQ Sequence 2733 AA;

Query Match 95.78; Score 1931; DB 22; Length 2733;
 Best Local Similarity 98.98; Pred. No. 2.3e-151;
 Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 10 KPSAEGRPPTPTSSPSLLPSAQLPSSHNPPVSCOMPLDLSNTSHQIMDTNDEEFSN 69
 DB 162 KSDDEGRPIPTTSSPSLLPSAQLPSSHNPPVSCOMPLDLSNTSHQIMDTNDEEFSN 221
 QY 70 SYLLRACSGQQASSSSGPPNHHQSOTLRPLPPHNHTLSHHSSANSLSNRSLTNRRSQ 129
 DB 222 SYLLRACSGQQASSSSGPPNHHQSOTLRPLPPHNHTLSHHSSANSLSNRSLTNRRSQ 281
 QY 130 IHAPAPNDLATTPESVQLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSPGYPLTS 189
 DB 282 IHAPAPNDLATTPESVQLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSPGYPLTS 341
 QY 190 GTVYTPPRLPLRNTFSKAFKLPKSKYCSWKCAALSAIAAALLAILLAYFTVPWSLK 249
 DB 342 GTVYTPPRLPLRNTFSKAFKLPKSKYCSWKCAALSAIAAALLAILLAYFTVPWSLK 401
 QY 250 NSSIDSGAEVGRVTOEVPVPGVFWRSQIHISQPFKFNISLGKDALFGVYIRRGPLPS 309
 DB 402 NSSIDSGAEVGRVTOEVPVPGVFWRSQIHISQPFKFNISLGKDALFGVYIRRGPLPS 461
 QY 310 HAQYDFMERLDGKEKWSVESPRRRRSIQTLVQNEAVFVQYLDVGLWHLAFYNDGDKDEM 369
 DB 462 HAQYDFMERLDGKEKWSVESPRRRRSIQTLVQNEAVFVQYLDVGLWHLAFYNDGDKDEM 521
 QY 370 VSFNTVVLD 378
 DB 522 VSFNTVVLD 530

RESULT 6

ABG03234
 ID ABG03234 standard; Protein; 865 AA.

XX
 AC ABG03234;

XX 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #3225.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS67421.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX

PS Claim 20; SEQ ID No 33593; 103pp; English.

XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 865 AA;

Query Match 49.28; Score 992.5; DB 22; Length 865;
 Best Local Similarity 61.88; Pred. No. 8.1e-74;
 Matches 214; Conservative 12; Mismatches 47; Indels 73; Gaps 6;

QY 84 SSGPPNHHSSOSTLRPLPPPHNHTLSHHSSANSLSNRSLTNRRSQIHAPAPNDLATTT 143
 DB 158 SSTAPSTPTTTRCPITTP-----SANSLSNRSLTNRRSQIHAPAPNDLATTT 207
 QY 144 PESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSPGYPLTSGTVYTPPRLPLRN 203
 DB 208 PESVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSPGYPLTSGTVYTPPRLPLRN 267
 QY 204 TFSKAFKLPKSKYCSWKCAALSAIAAALLAILLAYFTVPWSLKSSIDSGAE----- 259
 DB 268 TFSKAFKLPKSKYCSWKCAALSAIAAALLAILLAYFTIAGPQMERKASSQVKDRRLR 327
 QY 260 -----VGRRTVQEVPPGVFWRSQIHISQPFKFNISLGKDALFGVYIRRGILP----- 307
 DB 328 EILLFVGNSTVSPFPDP--WRQTV-----QEMMGKSKSCFMRHMTLMTWTAGI 375
 QY 308 -----PSH-----AOYDFMERLDGKEKWSVESPR 332
 DB 376 ISCILOKQOAWAGERTCPRHKASKQODLRITKKIIVDATYDFMERLDGKEKWSVESPR 435
 QY 333 ERRSIQTLVQNEAVFVQYLDVGLWHLAFYNDGDKEMVSEFNTVVLD 378
 DB 436 ERRSIQTLVQNEAVFVQYLDVGLWHLAFYNDGDKEMVSEFNTVVLD 481

RESULT 7

ABP53587

ID ABP53587 standard; Protein; 2721 AA.

XX
 AC ABP53587;

XX 17-DEC-2002 (first entry)

XX Human NOVI5b protein SEQ ID NO:38.

XX Human; NOVX; cytostatic; neuroprotective; anticonvulsant; cardiovascular;
 KW cerebroprotective; nootropic; antidiabetic; antiinflammatory; fungicide;
 KW antirheumatic; antiarthritic; immunosuppressive; antiallergic; virucide;
 KW antianaemic; antibacterial; protozoacide; antihelminthic; gene therapy;

cancer; leukaemia; lymphoma; melanoma; neurological disorder; epilepsy; stroke; ischaemic cerebrovascular disease; Alzheimer's disease; allergy; Pick's disease; vesicular transport disease; cystic fibrosis; goitre; diabetes mellitus; Grave's disease; gastrointestinal disorder; vaccine; ulcerative colitis; gastric disorder; duodenal disorder; infection; autoimmune disease; allergic reaction; autoimmune haemolytic anaemia; rheumatoid arthritis; chromosome 4.

Homo sapiens.

WO200262999-A2.

15-AUG-2002.

31-DEC-2001: 2001WO-US49976

29-DEC-2000: 2000TS-258928P

02-JAN-2001; 2001US-259415P.

20-FEB-2001; 2001US-269814P.

29-MAR-2001; 2001US-279832P.

29-MAR-2001; 2001US-279833P.
13-APR-2001: 2001US-283889P

18-APR-2001; 2001US-284447P.

29-MAY-2001; 2001US-294080P.

17-AUG-2001; 2001US-313325P.

17-SEP-2001: 20010S-322699P.
26-NOV-2001: 20010S-333350P

1. **Introduction**
 2. **Background**
 3. **Methodology**
 4. **Results**
 5. **Discussion**
 6. **Conclusion**
 7. **References**
 8. **Appendix**
 9. **Figure 1**
 10. **Figure 2**
 11. **Figure 3**
 12. **Figure 4**
 13. **Figure 5**
 14. **Figure 6**
 15. **Figure 7**
 16. **Figure 8**
 17. **Figure 9**
 18. **Figure 10**
 19. **Figure 11**
 20. **Figure 12**
 21. **Figure 13**
 22. **Figure 14**
 23. **Figure 15**
 24. **Figure 16**
 25. **Figure 17**
 26. **Figure 18**
 27. **Figure 19**
 28. **Figure 20**
 29. **Figure 21**
 30. **Figure 22**
 31. **Figure 23**
 32. **Figure 24**
 33. **Figure 25**
 34. **Figure 26**
 35. **Figure 27**
 36. **Figure 28**
 37. **Figure 29**
 38. **Figure 30**
 39. **Figure 31**
 40. **Figure 32**
 41. **Figure 33**
 42. **Figure 34**
 43. **Figure 35**
 44. **Figure 36**
 45. **Figure 37**
 46. **Figure 38**
 47. **Figure 39**
 48. **Figure 40**
 49. **Figure 41**
 50. **Figure 42**
 51. **Figure 43**
 52. **Figure 44**
 53. **Figure 45**
 54. **Figure 46**
 55. **Figure 47**
 56. **Figure 48**
 57. **Figure 49**
 58. **Figure 50**
 59. **Figure 51**
 60. **Figure 52**
 61. **Figure 53**
 62. **Figure 54**
 63. **Figure 55**
 64. **Figure 56**
 65. **Figure 57**
 66. **Figure 58**
 67. **Figure 59**
 68. **Figure 60**
 69. **Figure 61**
 70. **Figure 62**
 71. **Figure 63**
 72. **Figure 64**
 73. **Figure 65**
 74. **Figure 66**
 75. **Figure 67**
 76. **Figure 68**
 77. **Figure 69**
 78. **Figure 70**
 79. **Figure 71**
 80. **Figure 72**
 81. **Figure 73**
 82. **Figure 74**
 83. **Figure 75**
 84. **Figure 76**
 85. **Figure 77**
 86. **Figure 78**
 87. **Figure 79**
 88. **Figure 80**
 89. **Figure 81**
 90. **Figure 82**
 91. **Figure 83**
 92. **Figure 84**
 93. **Figure 85**
 94. **Figure 86**
 95. **Figure 87**
 96. **Figure 88**
 97. **Figure 89**
 98. **Figure 90**
 99. **Figure 91**
 100. **Figure 92**
 101. **Figure 93**
 102. **Figure 94**
 103. **Figure 95**
 104. **Figure 96**
 105. **Figure 97**
 106. **Figure 98**
 107. **Figure 99**
 108. **Figure 100**
 109. **Figure 101**
 110. **Figure 102**
 111. **Figure 103**
 112. **Figure 104**
 113. **Figure 105**
 114. **Figure 106**
 115. **Figure 107**
 116. **Figure 108**
 117. **Figure 109**
 118. **Figure 110**
 119. **Figure 111**
 120. **Figure 112**
 121. **Figure 113**
 122. **Figure 114**
 123. **Figure 115**
 124. **Figure 116**
 125. **Figure 117**
 126. **Figure 118**
 127. **Figure 119**
 128. **Figure 120**
 129. **Figure 121**
 130. **Figure 122**
 131. **Figure 123**
 132. **Figure 124**
 133. **Figure 125**
 134. **Figure 126**
 135. **Figure 127**
 136. **Figure 128**
 137. **Figure 129**
 138. **Figure 130**
 139. **Figure 131**
 140. **Figure 132**
 141. **Figure 133**
 142. **Figure 134**
 143. **Figure 135**
 144. **Figure 136**
 145. **Figure 137**
 146. **Figure 138**
 147. **Figure 139**
 148. **Figure 140**
 149. **Figure 141**
 150. **Figure 142**
 151. **Figure 143**
 152. **Figure 144**
 153. **Figure 145**
 154. **Figure 146**
 155. **Figure 147**
 156. **Figure 148**
 157. **Figure 149**
 158. **Figure 150**
 159. **Figure 151**
 160. **Figure 152**
 161. **Figure 153**
 162. **Figure 154**
 163. **Figure 155**
 164. **Figure 156**
 165. **Figure 157**
 166. **Figure 158**
 167. **Figure 159**
 168. **Figure 160**
 169. **Figure 161**
 170. **Figure 162**
 171. **Figure 163**
 172. **Figure 164**
 173. **Figure 165**
 174. **Figure 166**
 175. **Figure 167**
 176. **Figure 168**
 177. **Figure 169**
 178. **Figure 170**
 179. **Figure 171**
 180. **Figure 172**
 181. **Figure 173**
 182. **Figure 174**
 183. **Figure 175**
 184. **Figure 176**
 185. **Figure 177**
 186. **Figure 178**
 187. **Figure 179**
 188. **Figure 180**
 189. **Figure 181**
 190. **Figure 182**
 191. **Figure 183**
 192. **Figure 184**
 193. **Figure 185**
 194. **Figure 186**
 195. **Figure 187**
 196. **Figure 188**
 197. **Figure 189**
 198. **Figure 190**
 199. **Figure 191**
 200. **Figure 192**
 201. **Figure 193**
 202. **Figure 194**
 203. **Figure 195**
 204. **Figure 196**
 205. **Figure 197**
 206. **Figure 198**
 207. **Figure 199**
 208. **Figure 200**
 209. **Figure 201**
 210. **Figure 202**
 211. **Figure 203**
 212. **Figure 204**
 213. **Figure 205**
 214. **Figure 206**
 215. **Figure 207**
 216. **Figure 208**
 217. **Figure 209**

1997

Malyankar U, Shimkets RA,

Kekuda R, Patturajan M, Gu
Rastelli L, Casman S.T. Bol

Ellerman K, Gunther E, Smi

WPI; 2002-732706/79.

(CURA-) CURAGEN CORP.

Spytek KA, Li L, Wolenc AR, Vernet CAM, Elsen A, Liu X;

Malyankar U, Shimkets RA, Tchernev VT, Spaderna SK, Gorman L;

Kekuda R, Patturajan M, Gusev V, Gangoli EA, Guo X, Shenoy S;
Rastelli L; Casman SJ. Boldog F, Burgess CE, Edinger S.

Ellerman K, Gunther E, Smithson G, Millet I, MacDougall JR;

C
E
V
C
F
C
Z

C
C
C
C

F
S
S

WPI; 2002-732706/79.
N-PSDB: ABQ82344

[illegible]

New NOVX polypeptides and polynucleotides useful for treating

NOVX-associated disorders, such as cancers, neurological disorders and disorders of vesicular transport gastrointestinal disorders and

autoimmune diseases -

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Claim 1; Page 117-118; 444pp; English

The present invention describes novel human proteins designated NOVX, where X is 1 to 20 e.g. NOV1. NOVX sequences can have neuroprotective, cytostatic, anticonvulsant, cerebroprotective, nootropic, cardiovascular, antidiabetic, antiinflammatory, antirheumatic, antiarthritic, virucide, immunosuppressive, antifungal, antianemic, antibacterial, fungicide, protozoicide and antihelminthic activities, and can be used in gene therapy. The NOVX proteins, nucleotides or antibodies can be used in the manufacture of a medicament for treating a syndrome associated with a human disease selected from NOVX-associated disorder, such as cancers (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle, ovary, testis and uterus), neurological disorders (e.g. epilepsy, stroke, ischaemic cerebrovascular disease, Alzheimer's disease or Pick's disease), disorders of vesicular transport (e.g. cystic fibrosis, diabetes mellitus, Grave's disease, or goitre), gastrointestinal disorders (e.g. ulcerative colitis, or gastric and duodenal allergic disorders), autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic and protozoal infections. The NOVX proteins can be used as immunogens to produce antibodies and as vaccines. The NOVX nucleotide sequences may be used in chromosome mapping, identifying individuals from minute biological samples (tissue typing), and in forensic identification of a biological sample. The present sequence represents human NOV15b, which is located on chromosome 4.

CC The present invention describes novel human proteins designated NOVX,
 CC where X is 1 to 20 e.g. NOV1. NOVX sequences can have neuroprotective,
 CC cytostatic, anticonvulsant, cerebroprotective, neurotropic, cardiovascular,
 CC antidiabetic, antiinflammatory, antirheumatic, antiarthritic, virucide,
 CC immunosuppressive, antiallergic, antianaemic, antibacterial, fungicide,
 CC protozoacide and antihelminthic activities, and can be used in gene
 CC therapy. The NOVX proteins, nucleotides or antibodies can be used in the
 CC manufacture of a medicament for treating a syndrome associated with a
 CC human disease selected from NOVX-associated disorder, such as cancers
 CC (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,
 CC ovary, testis and uterus), neurological disorders (e.g. epilepsy, stroke,
 CC ischaemic cerebrovascular disease, Alzheimer's disease or Pick's
 CC disease), disorders of vesicular transport (e.g. cystic fibrosis,
 CC diabetes mellitus, Grave's disease, or goitre), gastrointestinal
 CC disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),
 CC autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic
 CC anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic
 CC and protozoal infections. The NOVX proteins can be used as immunogens to
 CC produce antibodies and as vaccines. The NOVX nucleotide sequences may
 CC be used in chromosome mapping, identifying individuals from minute
 CC biological samples (tissue typing), and in forensic identification of a
 CC biological sample. The present sequence represents human NOV15c, which is
 CC located on chromosome 4.
 CC
 CC Sequence 2628 AA;

Query Match 48.2%; Score 972.5; DB 23; Length 2628;

Best Local Similarity 55.1%; Pred. No. 1.7e-71;

Matches 190; Conservative 46; Mismatches 58; Indels 51; Gaps 5;

QY 87 PNNHSSQVLRPLPPPHNHTLSHHSSANSLSNRSLNRSQIHA-PAPAPNDLATPTE 145
 DB 80 PASNOGQSTLQ-PLPPSHKQHSQAHHPSITSLNRSLNRRNQSPAPPALPAELQTTPTE 138
 QY 146 SVQLQDSWVLSNVPLETRHFLFKTSSTGSGTPLFSSSSPGVPLTSGTVTPPPRLPRNTF 205
 DB 139 SVQLQDSWVLSNVPLESRHFLEKTKTGTPTPLFSTATPGTVMASGVSVPPTPLPRNTL 198
 QY 206 SRKAFKKPKSCYKWCALSAIAAALLAILLAYFI-----VPSWL----- 248
 DB 199 SRSAFKKSKYKWCALCAVGVSVLLAILLSYFIAMHLFGLNWLQQTENDTFENG 258
 QY 249 -----KNSSDSGEAEVGRVTVQVPPGVFWRWSQIHIS 291
 DB 259 KVNSDTMTNTVSLPSGDKLGFGFTQENNTDSGELDIGRAIOIPPGIFWRWSQLFID 318
 QY 282 QPQFLKFNISLQKDALFYVIRRGPLPSHAQYDFMERLDG-----KEKSVVESPRERS 336
 DB 319 QPQFLKFNISLQKDALIGYGRKGLPPSHITQDFVELLDGSLRIAREQSLLETERAGRQ 378
 QY 337 IOTLVQNEAVFYQYLDVGLWHLAFYNDGDKEMVSFNTVYLDGTI 381
 DB 379 ARSVSLHEAGFIQYLDGSIWHLAFYNDGKNAEQVSNTIVIESVW 423

RESULT 10

ABP53586

ID 'ABP53586 standard; Protein; 2725 AA.

XX AC ABP53586;

XX AC ABP53586;

DT 17-DEC-2002 (first entry)

XX Human NOV15a protein SEQ ID NO:36.

XX Human: NOVX: cytostatic; neuroprotective; anticonvulsant; cardiovascular;
 KW cerebroprotective; neurotropic; antidiabetic; antiinflammatory; fungicide;
 KW antirheumatic; antiarthritic; immunosuppressive; antiallergic; virucide;
 KW antianaemic; antibacterial; protozoacide; antihelminthic; gene therapy;
 KW cancer; leukaemia; lymphoma; melanoma; neurological disorder; epilepsy;
 KW stroke; ischaemic cerebrovascular disease; Alzheimer's disease; allergy;
 KW Pick's disease; vesicular transport disease; cystic fibrosis; goitre;
 KW diabetes mellitus; Grave's disease; gastrointestinal disorder; vaccine;

KW ulcerative colitis; gastric disorder; duodenal disorder; infection;
 KW autoimmune disease; allergic reaction; autoimmune haemolytic anaemia;
 XX rheumatoid arthritis; chromosome 4.
 OS Homo sapiens.
 XX WO200262999-A2.
 XX 15-AUG-2002.

XX 31-DEC-2001; 2001WO-US49976.

XX 29-DEC-2000; 2000US-258928P.

XX 02-JAN-2001; 2001US-259415P.

XX 04-JAN-2001; 2001US-259785P.

XX 20-FEB-2001; 2001US-269814P.

XX 09-MAR-2001; 2001US-279863P.

XX 29-MAR-2001; 2001US-279832P.

XX 29-MAR-2001; 2001US-279833P.

XX 13-APR-2001; 2001US-283889P.

XX 18-APR-2001; 2001US-284447P.

XX 25-APR-2001; 2001US-286683P.

XX 29-MAY-2001; 2001US-294080P.

XX 16-AUG-2001; 2001US-312915P.

XX 17-AUG-2001; 2001US-31325P.

XX 17-SEP-2001; 2001US-322699P.

XX 26-NOV-2001; 2001US-333350P.

XX (CURA-) CURAGEN CORP.

XX SPYTEK KA, LI L, Wolenc AR, Vernet CAM, Eisen A, Liu X;

XX Malyankar U, Shinkets RA, Tchernev VT, Spaderna SK, Gorman L;

XX Kekuda R, Patturajan M, Gusev V, Gangolli EA, Guo X, Shenoy S;

XX Rastelli L, Casman SJ, Boldog F, Burgess CE, Edinger S;

XX Ellerman K, Gunther E, Smithson G, Millet I, MacDougall JR;

XX WPI; 2002-732706/79.

XX N-PSDB; ABQ82343.

XX New NOVX polypeptides and polynucleotides useful for treating

XX NOVX-associated disorders, such as cancers, neurological disorders,

XX disorders of vesicular transport, gastrointestinal disorders, and

XX autoimmune diseases -

XX Claim 1; Page 113; 444pp; English.

XX The present invention describes novel human proteins designated NOVX,

XX where X is 1 to 20 e.g. NOV1. NOVX sequences can have neuroprotective,

XX cytostatic, anticonvulsant, cerebroprotective, neurotropic, cardiovascular,

XX antidiabetic, antiinflammatory, antirheumatic, antiarthritic, virucide,

XX immunosuppressive, antiallergic, antianaemic, antibacterial, fungicide,

XX protozoacide and antihelminthic activities, and can be used in gene

XX therapy. The NOVX proteins, nucleotides or antibodies can be used in the

XX manufacture of a medicament for treating a syndrome associated with a

XX human disease selected from NOVX-associated disorder, such as cancers

XX (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,

XX ovary, testis and uterus), neurological disorders (e.g. epilepsy, stroke,

XX ischaemic cerebrovascular disease, Alzheimer's disease or Pick's

XX disease), disorders of vesicular transport (e.g. cystic fibrosis,

XX diabetes mellitus, Grave's disease, or goitre), gastrointestinal

XX disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),

XX autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic

XX anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic

XX and protozoal infections. The NOVX proteins can be used as immunogens to

XX produce antibodies and as vaccines. The NOVX nucleotide sequences may

XX be used in chromosome mapping, identifying individuals from minute

XX biological samples (tissue typing), and in forensic identification of a

XX biological sample. The present sequence represents human NOV15a, which is

XX located on chromosome 4.
 XX
 XX Sequence 2725 AA;

Query Match 47.4%; Score 956.5; DB 23; Length 2725;


```
Best Local Similarity 50.18; Pred. No. 3.8e-70;
Matches 193; Conservative 53; Mismatches 70; Indels 69; Gaps 7;

Qy 51 SNTSHQIMDYNPDEFSNPSYLLRACSGPQOASSGPPNHHSSQTLRPPPLPPPHNHTLSH 110
   ||: : ||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 SNSALTLDTEHKN-----SDSENEQPASNOGSTLQ-PLPPSHKHQSAQ 195
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :

Qy 111 HHSANSLNRLNRRSQTHA-PAPAPNDLATTPEVSQVLODSWVLNSNVPLETRHFLFK 169
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 196 HHSITSLNRLNRRNQSPAPPALPAELQTTPEVSQVLODSWVLNSNVPLESRHFLFK 255
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :

Qy 170 TSSGSTPLFSSSPCYPLTSGTYTPPPRLPLRNTFSRKAPKPKSKYCSWKCAALSAI 229
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 256 TGTGTTPLFATVATPGYTWAGSVTPTRPLRNLRSAPFKFKSKYCSWKCTALCAV 315
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :

Qy 230 AAALLAILLAYFI-----VPWSL----- 248
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 316 GSVLLAILLSYFTAMHLFGLNWLOQTENDTPENGKVNSTMTNTVSLPSGDNGKLG 375
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :

Qy 249 ---KNSIDSGEAEVGRVTVQEVPPGVFWRSQIHISQPOFLKFNISLKGKDALFGVYTRG 305
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 376 FTQENNTIDSGELDIGRAIQEIPPGIFWRSOLFIDQPFKFNISLQKDALIGVYGRKK 435
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :

Qy 306 LPPSHA---QYDPMERLDG-----KEKWSVSPRRRSITQTLVQNEAVFVQYLDVGLW 356
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 436 LPPSHTOSSQYDVELLDSRLIARQSRLLTERAGRAQRSVSLHEAGFIQYLDVGLW 495
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :

Qy 357 HLAIFYNDGDKEMVSFNTVVDGTI 381
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 496 HLAIFYNDGKNAEQVSFNTIVIESVV 520
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :

RESULT 11
ABG97359
ID ABG97359 standard; Protein: 2758 AA.
XX AC ABG97359;
XX DT 16-DEC-2002 (first entry)
XX DE Human CGDD10, INCYTE 7488573CD1.
XX KW Human; cell growth; differentiation; death; CGDD; cancer;
XX KW cell proliferative disorder; arteriosclerosis; atherosclerosis;
XX KW cirrhosis; hepatitis; paroxysmal nocturnal haemoglobinuria; psoriasis;
XX KW polycythaemia vera; primary thrombocytopaenia; developmental disorder;
XX KW renal tubular acidosis; anaemia; mental retardation; epilepsy; AIDS;
XX KW neurological disorder; Alzheimer disease; Parkinson's disease; asthma;
XX KW reproductive disorder; infertility; autoimmune disorder; gout; allergy;
XX KW inflammatory disorder; acquired immunodeficiency syndrome; uveitis;
XX KW autoimmune thyroiditis; contact dermatitis; Crohn's disease; infection;
XX KW diabetes mellitus; glomerulonephritis; irritable bowel syndrome;
XX KW multiple sclerosis; osteoarthritis; osteoporosis; pancreatitis;
XX KW rheumatoid arthritis.
XX OS Homo sapiens.
XX FN WO200272830-A2.
XX PD 19-SEP-2002.
XX PF 08-FEB-2002; 2002WO-US03715.
XX PR 09-FEB-2001; 2001US-268111P.
XX PR 23-FEB-2001; 2001US-271175P.
XX PR 08-MAR-2001; 2001US-274503P.
XX PR 09-MAR-2001; 2001US-274552P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Yue H, Yao MG, Ison CH, Lu Y, Warren BA, Elliott VS, Baughn MR;
PI Ding L, Xu Y, Gietzen KJ, Tang TY, Lal PG, Duggan BW, Burford N;
PI Lu DAW, Richardson TW, Tran UK, Khare R, Wallia NK;
```

XX WPI; 2002-723356/78.
DR N-PSDB; ABS78652.

XX New human proteins associated with cell growth, differentiation and
death, useful for diagnosing, treating or preventing autoimmune or
inflammatory disorders (e.g. AIDS, allergy or anemia), cancer,
atherosclerosis or hepatitis
XX Claim 1; Page 155-161; 181pp; English.

XX The invention relates to an isolated polypeptide comprising CGDD1-12
(cell growth, differentiation and death), a naturally occurring amino
acid sequence at least 90% identical to CGDD, a biologically active
fragment or an immunogenic fragment. Also included are the
polynucleotides encoding CGDD1-12, a recombinant polynucleotide
comprising a promoter sequence operably linked to the CGDD
polynucleotides, a cell transformed with the recombinant polynucleotide,
a transgenic organism comprising the recombinant polynucleotide, an
anti-CGDD antibody, screening for compounds which bind to/modulate
or are ant/agonists of CGDD or alter the expression of CGDD
polynucleotide and a CGDD polynucleotide microarray.
XX The polypeptides, polynucleotides, agonists and antagonists are
useful for diagnosing, treating or preventing disorders associated with
aberrant expression of CGDD, particularly cell proliferative (e.g.
arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, paroxysmal
nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary
thrombocytopaenia or cancer), developmental disorders (e.g. renal
tubular acidosis, anaemia or mental retardation), neurological disorders
(e.g. Alzheimer disease, Parkinson's disease or epilepsy),
reproductive disorders (e.g. infertility or a disruption in the
menstrual cycle), or autoimmune/inflammatory disorders (e.g. AIDS,
(acquired immunodeficiency syndrome) allergies, asthma, autoimmune
thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus,
glomerulonephritis, gout, irritable bowel syndrome, multiple sclerosis,
osteoarthritis, osteoporosis, pancreatitis, rheumatoid arthritis,
uveitis, or viral, bacterial, fungal, parasitic, protozoal or helminthic
infections. They are also useful in the assessment of the effects of
exogenous compounds on the expression of nucleic acid and amino acid
sequences of proteins associated with CGDD. The present sequence
XX represents a CGDD protein.

XX Sequence 2758 AA;

Query Match 39.3%; Score 793; DB 23; Length 2758;
Best Local Similarity 40.5%; Pred. No. 1.6e-56;
Matches 179; Conservative 49; Mismatches 82; Indels 132; Gaps 13;

Qy 43 SCQPLDLSNTSHOIMDTNPDEFSNPSYLLRACSGPQOASSGPPNHHSSQTLRPPLP 102
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 142 SCLSSRANSNLT--LTDTEHENTETP-----GGLOQH---ARLRTPPPP 180
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Qy 103 -PHNHTLSHHH--SSANSLNRLNRRSQIHAPAPND--LATTP-----ESVQLQDS 152
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 181 LSHAHTPNQHHAAASINLNRGNFTPRSN----PSPAPTDHSLGSEPPAGGAQEAHAQEN 236
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Qy 153 WVLNSNVPLETR-----HFLKTSSTGSLPLFS 179
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 237 WLLNSNIPLETRNLGKQPFGLTLDNLEMDILGASRDGAYSDGHFLFK-PGTSPPLFC 295
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Qy 180 SSPGYPILTSCTVYTPPPRLPLRNTFSRKAPKPKSKYCSWKCAALSAIAALLL 239
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 296 TTSPGYPLTSTVYSPPPRPLRSTFARPAFNKPKSKYCNWKAALSAIVISATLVLL 355
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Qy 240 AYFI---VPWSL----- 248
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 356 AYFVKHLLFNHQLQPMQMEQVEITEDTASSWPFVPTDVSLSVPSGTGLTPDRKKGKGTG 415
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Qy 249 -----KNSIDSGEAEVGRVTVQEVPPGVFWRSQIHISQPOFLKFNISLKGKDALFGVY 301
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 416 KPSFFFEDESIDGSDVGRASQKPPGTFWRSQVFDHPVHLKFNWVSLGKALVGIY 475
 ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Qy 302 IRRGLPPSHAQYDFMERLDGK-----EKWSVSPRRRSITQTLVQNEAVFVQYLDVGLW 356

```

Db      476 GRKGLPSTHTQDFVLLDGRLLTQEARSLGTPQSRGTVPSPSSHERGFIQYLDGSIW 535
QY      357 HLAIFYNDGDKKEMVSFNTVVLID 378
Db      536 HLAIFYNDGKESEVVSFLTIAIE 557

RESULT 12
ABG70388
ID      ABG70388 standard; Protein; 2769 AA.
XX
AC      ABG70388;
XX
DT      05-NOV-2002 (first entry)
XX
DE      Human TEN-M4-like protein.
XX
KW      Human; NOVX; NOVX-associated disorder; cardiomyopathy; atherosclerosis;
KW      cell signal processing; metabolic pathway modulation; metabolic disorder;
KW      obesity; diabetes; infectious disease; neurodegenerative disorder; acne;
KW      Alzheimer's disease; Parkinson's disease; immune disorder; cancer;
KW      haematopoietic disorder; cirrhosis; pancreatitis; learning defect;
KW      memory defect; infertility; congenital heart defect; hair growth;
KW      pigmentation disorder; endocrine disorder; respiratory disease; health;
KW      gastro-intestinal disease; reproductive; neurological disease;
KW      bone marrow transplantation; endocrine disease; allergy; inflammation;
KW      neurological disorder; urinary system disorder; age-related disorder;
KW      neuropsychiatric disorder; EGF-related protein; SCUBEL; TEN-M4;
KW      adipocyte complement-related C1q tumour necrosis factor; out at first;
KW      beta adrenergic receptor kinase; EphA6/ehk-2; glucose transporter;
KW      type Ia membrane sushi-containing domain; butyrophilin;
KW      type Ia membrane-sushi domain containing; SNP;
KW      single nucleotide polymorphism.
XX
OS      Homo sapiens.
XX
FH      Key Location/Qualifiers
FT      Misc-difference 28 /note= "Asp substituted by Gly as a result of a single
FT      nucleotide polymorphism (SNP)"
FT      Misc-difference 64 /note= "Val substituted by Ala as a result of a single
FT      nucleotide polymorphism (SNP)"
FT      Misc-difference 76 /note= "Ala substituted by Thr as a result of a single
FT      nucleotide polymorphism (SNP)"
FT      WO200257453-A2.
XX
PD      25-JUL-2002.
XX
PF      19-DEC-2001; 2001WO-US50331.
XX
PR      19-DEC-2000; 2000US-265704P.
XX
PR      20-DEC-2000; 2000US-257314P.
XX
PR      02-MAY-2001; 2001US-288153P.
XX
PR      29-MAY-2001; 2001US-294075P.
XX
PR      24-JUL-2001; 2001US-307506P.
XX
PR      10-AUG-2001; 2001US-311590P.
XX
PR      10-AUG-2001; 2001US-311613P.
XX
PR      29-AUG-2001; 2001US-315617P.
XX
PR      14-SEP-2001; 2001US-322358P.
XX
PA      (CURA-) CURAGEN CORP.
XX
PI      Gangolli EA, Patturajan M, Vernet CAM, Malyankar UM, Kekuda R;
PI      Stone DJ, Anderson D, Shimkets RA, Burgess CE, Zerhusen BD, Liu X;
PI      Spytek KA, Casman SJ, Boldog FL, Smithson G, Li L, Ji W;
XX
DR      WPI; 2002-590744/63.
DR      N-PSDB; ABS52100.
XX

```

```

PT      Novel isolated NOVX polypeptide useful for treating cardiomyopathy,
PT      atherosclerosis, metabolic disorders, diabetes, obesity, infectious
PT      disease, anorexia, neurodegenerative disorders, Alzheimer's disease or
XX      cancer
XX
PS      Claim 1; Page 53; 318pp; English.
XX
CC      The present invention relates to new NOVX polypeptides. The invention is
CC      useful for treating or preventing a NOVX-associated disorder such as
CC      cardiomyopathy or atherosclerosis, where the disorder is related to cell
CC      signal processing and metabolic pathway modulation in a subject,
CC      preferably human. The invention is also useful for treating metabolic
CC      disorders (e.g. obesity), diabetes, infectious disease, neurodegenerative
CC      disorders (e.g. Alzheimer's disease, Parkinson's disease), immune
CC      disorders, haematopoietic disorders and various cancers. The molecules of
CC      the invention are also useful for treating or preventing cirrhosis, heart
CC      defects, acne, hair growth, pigmentation disorders, infertility, congenital heart
CC      respiratory disease, gastro-intestinal diseases, reproductive, health,
CC      neurological diseases, bone marrow transplantation, endocrine diseases,
CC      allergy and inflammation, nephrological disorders, urinary system
CC      disorders, neuropsychiatric disorders and age-related disorders.
CC      The present amino acid sequence represents a NOVX protein of the
CC      invention.
XX
SQ      Sequence 2769 AA;
XX
Query Match 39.3%; Score 792.5; DB 23; Length 2769;
Best Local Similarity 41.4%; Pred. No. 1.8e-56;
Matches 173; Conservative 49; Mismatches 83; Indels 113; Gaps 10;
QY      69 NSYLLRACSGPQOASSGPPNHHSSQSTLRPPPP-PHNHTLSHHH-SSANSLNRNLSLNR 126
Db      149 NSNLTLDTEHENTETDHPGGLQNHARLRTPPPLSHHTPNQHHAAASINSLNRGFTPR 208
QY      127 RSQIHAPAPND--LATTP-----ESVOLQDSWVLNSNVPLETR----- 164
Db      209 SN-----PSPAPTDHSLSGEPAGGAQOEPAAHQENWLLNSNIPLETRNLGKQPPLGTLQDN 264
QY      165 -----HFLFKTSSGSTPLFSSSGVPLTSGTVYTPPPRLPRNTF 205
Db      265 LIEMDILGASRDGAYSDGHFLFK-PGHSPLFCTTSPGTLTSSVTPPPRPLPRSTF 323
QY      206 SRKAFKKPKSCYKWCKAALSAIAAALLAILLAYFI-----VPWSL----- 248
Db      324 ARPAFLNKKPKSCYKWCKAALSAIVISATLVLLAYFVAMHFLGLNWHLQPMGQWYEIT 383
QY      249 -----KNSIDSIGEAEGVRRVT 265
Db      384 EDTASSWPVPTDVSLYPSGGTGLETPDRKGKGTCKSPSSFFPEDSFDSGETDVGRRAS 443
QY      266 QEVPPGVFWRISOIHISQPOFLKFNISLGDALFGVYVIRGLPPSHAOYDFMERLDGK--- 322
Db      444 QKIPGCTFWRSQVFDHPVHLAFNLSLGAALVGITGRKGLPPSHQTFQDFVELLDGRLL 503
QY      323 --EKWSVSPERRRSIQTLVQNEAVFYQYLDGLVHLLAFYNDGDKKEMVSFNTVVLID 378
Db      504 TQEARSLGTPQSRGTVPSPSSHERGFIQYLDGSIWHLAFYNDGKESEVVSFLTIAIE 561

RESULT 13
AAW44817
ID      AAW44817 standard; Protein; 768 AA.
XX
AC      AAW44817;
XX
DT      17-AUG-1998 (first entry)
XX
DE      Human gamma-heregulin.
XX
KW      Gamma-heregulin; gamma-HRG; human; autocrine growth factor;
KW      breast cancer; MDA-MB-175 cell; diagnosis; gene therapy;
KW      ErbB receptor; cell proliferation; cell differentiation;

```

cell survival; neurological disorder; muscular disorder.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX Domain 1..560

XX Region 342..363

FT /note= "N-terminal domain (Claim 5)"

FT /note= "hydrophobic region, may function as internal signal sequence"

FT Domain 704..748

FT /note= "EGF-like domain"

FT Modified-site 77

FT /note= "N-glycosylated"

FT Modified-site 151

FT /note= "N-glycosylated"

FT Modified-site 467

FT /note= "N-glycosylated"

FT Modified-site 647

FT /note= "N-glycosylated"

FT Modified-site 653

FT /note= "N-glycosylated"

FT Modified-site 691

FT /note= "N-glycosylated"

XX W09802541-A1.

XX 22-JAN-1998.

XX 08-JUL-1997; 97WO-US11841.

XX 12-JUL-1996; 96US-0021640.

XX (GETH) GENENTECH INC.

XX Schaefer GM, Sliwkowski M;

XX WPI; 1998-110589/10.

XX N-PSDB; AAV19251.

XX DNA encoding gamma-heretulin - used to activate ErbB receptor and to enhance proliferation, differentiation or survival of a cell

XX Claim 3; Fig 1A-C; 81pp; English.

XX This polypeptide comprises human gamma-heretulin (gamma-HRG), a novel member of the heretulin superfamily, that has a unique N-terminal domain not present in previously identified heregulins. Gamma-HRG is secreted by human breast cancer MDA-MB-175 cells. It leads to the formation of a constitutive active receptor complex and stimulates the growth of these cells in an autocrine manner. A claimed method for activating an ErbB receptor comprises contact a cell which expresses the receptor with gamma-HRG. A claimed method for enhancing the proliferation, differentiation or survival of a cell, especially a glial cell or muscle, comprises contacting the cell with gamma-HRG. Antibodies raised against the unique N-terminal domain of gamma-HRG can be used to detect and purify the protein. Antagonists of gamma-HRG can be used to block gamma-HRG activity and expression. The gamma-HRG may be obtained from claimed host cells that comprise a vector containing an isolated gamma HRG nucleic acid (see AAV19251).

XX Sequence 768 AA;

Query Match 39.2%; Score 790.5; DB 19; Length 768;

Best Local Similarity 41.5%; Pred. No. 4.7e-57;

Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;

QY 69 NSYLLRACSGPQQASSGPPHHQSQTLPPLPP-PHNHTLSHHH-SSANSINRSLNTR 126

DB 149 NSNLTLDTEHENTETDRPGGLQNHARLRTPPPLPSHAHTPNQHHAAASINRNCNFTPR 208

QY 127 RSQIHAPAPND--LATP-----ESVQLQDSWLVNSNPVLETR----- 164

DB 209 SN-----PSPAPTDHSLSGEPPAGGAQEPAAHQENLLNSNIPLETRNLGKQPLGTLQDN 264

QY 165 -----HLEKTSSTSGSTPLFSSSSPGYPLTSTVTPPPRLPRNTE 205

DB 265 LIEMDILGASRDGAYSDGHLFK-PGTSPLFCTTSPGYPLTSTVTPPPRLPRSTF 323

QY 206 SRKAFKLKPKSKYCSWKCAALSAIAAALLAILLAYFI-----VPWSL----- 248

DB 324 ARPANLKKPKSKYCNWKAALSAIVISATLVILLAYFVAMHLFGLNHLQPMWEGOMYEIT 383

QY 249 -----KNSSTDSEAEVGRRTV 265

DB 384 EDTASSNPVPTDVSILYPSGGTGLETPDRKKGTTGKPSFFPEDSFIDSGEIDVGRRAS 443

QY 266 QEVPPGVFWRQIHSIQPOFLKFNISLGKDALFGVYIRRGLPSPHAQYDFMERLDGK--- 322

DB 444 QKIPPGTWRQSVQFDHPVHLKFNVSLGKAALVIGYGRKGLPPSTQTFDFVELLDGRLL 503

QY 323 --EKWSVVESPRRRSIQTLVONEAVFYQYLDVGLWHLAFYNDGKDEKEMVSFNTVYL 377

DB 504 TQEARSLGTPRQSRGTVPSPSHETGFIQYLDGSIWHLAFYNDGKESEWSFLTTAI 560

RESULT 14

AA06639

ID AAY06639 standard; Protein; 768 AA.

XX AC AAY06639;

XX DT 26-OCT-1999 (first entry)

XX DE Gamma-heretulin.

XX KW Gamma-heretulin; HRG-alpha; human; ligand; HER2; HER3; HER4;

XX KW receptor; lung surfactant; respiratory distress syndrome;

XX KW emphysema; epithelial growth factor; therapy.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX Region 342..363

FT Modified-site 77

FT /note= "hydrophobic region"

FT Modified-site 151

FT /note= "N-glycosylated"

FT Modified-site 467

FT /note= "N-glycosylated"

FT Modified-site 647

FT /note= "N-glycosylated"

FT Modified-site 653

FT /note= "N-glycosylated"

FT Modified-site 691

FT /note= "N-glycosylated"

FT Domain 704..748

FT /note= "epidermal growth factor-like domain"

XX W09939729-A2.

XX 12-AUG-1999.

XX 03-FEB-1999; 99WO-US02390.

XX 04-FEB-1998; 98US-0020598.

XX (GETH) GENENTECH INC.

XX PA (IOWA) UNIV IOWA RES FOUND.

XX PI Kern JA, Sliwkowski M;

XX WPI; 1999-494213/41.

XX N-PSDB; AAX87705.

XX Heregulin ligands can be used to induce epithelial cell growth, and
PT to promote repair and healing of tissue damage or injury
XX
PS
XX
XX Disclosure; Page 108-111; 120pp; English.
XX This sequence represents gamma-hergulin (gamma-HRG) deduced from
CC gamma-HRG cDNA (see AAX87705). The invention provides HRG ligands,
CC including gamma-HRG, that have affinity for and stimulate HER2,
CC HER3 and/or HER4 receptors in autophosphorylation. A new method of
CC treating respiratory distress syndrome in humans uses HER2, HER3
CC and/or HER4 receptor ligands as epithelial growth factors. A novel
CC method of inducing epithelial cell growth and/or proliferation
CC comprises contacting a normal epithelial cell which expresses HER2,
CC HER3 and/or HER4 receptors with an isolated ligand which activates
CC HER2, HER3, HER4 receptors or their combination. Also claimed are
CC methods of increasing lung surfactant protein A, or of treating
CC chronic obstructive pulmonary disease, respiratory distress or
CC emphysema, by administering an effective amount of an isolated HER
CC ligand to a patient.
XX
XX Sequence 768 AA;
SQ
Query Match 39.2%; Score 790.5; DB 20; Length 768;
Best Local Similarity 41.5%; Pred. No. 4.7e-57;
Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;
QY 69 NSYLLRACSGPOQASSGPPNHHSSQSTLRPPPLPP-PHNHTLSHHH-SSANSLNRLNLTNR 126
Db 149 NSNLTLDTEHENTETDHPGGLQNHARLTTPPPPLSHAHTPNQHHAAASINSLNRGNFTPR 208
QY 127 RQSIHAPAPND--LATTP-----ESVOLQDSWVLNSNVPLETR----- 164
Db 209 SN-----PSPAPT DSHLSGEPAGGAGQEPAAHQAENWLLNSNIPLETNRNKGKQPLGLTQDN 264
QY 165 -----HFLFKTSSGTPLEFSSSSPGYPLTSGTVYTPPPRLPRNTF 205
265 LEMDILGASRDGAYSDGHEFLK-PGGTSPLFCFTSPGTYLTSSTVSPPPRPLPRSTF 323
QY 206 SRKAFKLRKPSYCWKCAALSAAIAAALLLAILAYFI-----VPWSL----- 248
Db 324 ARPANLKRKPSYCWKCAALSAAIAVISATLVILLAYFVAMHLFGLNHLQPMQGYEIT 383
QY 249 -----KNSIDSGBAEVGRVY 265
Db 384 EDTASSWPVPTDVSLYPSGGTGLETPDRKGKGTETGKPSFFPEDSFIDSGEIDVGRAS 443
QY 266 QEVPPGVFWRSHISQPOFLFNISLGDALFVYIRGLPPLSHAQYDFMERLDGK--- 322
Db 444 QKIPPGTFWRQSVFIDHPVHLKFNYSGLKAAVLGYIGRKGPLPSPHTQDFVELLDGRLL 503
QY 323 --EKWSVVPSPRRRSIQTLQNEAVFVOYLDVGLWHLAFYNDGDKEMVSENTVVL 377
Db 504 TQEARSLGTPQSRGTPPPSSHVGFQIQLDSGWLHAFYNDGRESEVSLTAT 560
RESULT 15
AAV71176
ID AAV71176 standard; Protein; 768 AA.
XX
AC AAV71176;
XX
XX 21-SEP-2000 (first entry)
XX
XX Human Heregulin variant, gamma-HRG protein.
DE
KW Heregulin; variant; gamma-HRG; human; inner-ear-supporting cell;
KW activator; HER2 receptor; HER3 receptor; HER-2 oncogene; chromosome 8p;
KW hair cell related hearing disorder; ototoxic injury; tissue damage;
KW acoustic assault; degenerative hearing loss; balance impairment;
KW treatment; surgical injury; physical injury; inner ear disorder.
XX
XX Homo sapiens.
OS

XX Key Location/Qualifiers
FH Modified-site 77
FT /note= "Asn is N-glycosylated"
XX Modified-site 151
FT /note= "Asn is N-glycosylated"
XX Region 342..363
FT /label= Hydrophobic_region
FT Modified-site 467
FT /note= "Asn is N-glycosylated"
FT Modified-site 647
FT /note= "Asn is N-glycosylated"
FT Modified-site 653
FT /note= "Asn is N-glycosylated"
FT Modified-site 691
FT /note= "Asn is N-glycosylated"
FT Domain 704..748
FT /label= EGF_like_domain
FT /note= "Consists of six cysteine residues"
XX
PN WO200027426-A1.
XX
XX 18-MAY-2000.
XX
XX 28-OCT-1999; 99WO-US25744.
XX
XX 07-NOV-1998; 98US-0107522.
XX (GETH) GENENTECH INC.
XX Gao W;
XX
XX WPI; 2000-376313/32.
DR N-PSDB; AAD00791.
XX
XX Method for inducing hair cell generation and inner-ear-supporting cell
PT growth regeneration and proliferation, useful for treating hearing
PT disorders
XX
XX Disclosure; Fig 7; 141pp; English.
XX
XX The patent discloses a method for inducing hair cell generation, or
CC inner-ear-supporting cell growth, regeneration, and/or proliferation, by
CC heregulins (HRG), the ligand for HER2/HER3 dimeric receptors. Heregulin
CC proteins function as activators of HER-2 oncogene and result from
CC alternate splicing of a single gene mapped to chromosome 8p. The two
CC major types, alpha and beta HRG's are based on two variant EGF-like
CC (epidermal growth factor) domains, that differ in their C-terminal ends.
CC This method can be used to increase the number of inner-ear-supporting
CC cells and for treatment of hair cell related hearing disorders and
CC disease states associated with tissue damage, e.g. ototoxic injury,
CC acoustic assault, degenerative hearing loss, balance impairments, damage
CC associated with surgery or physical injury and inner ear disorders
CC related to hair cell dysfunction. The present sequence is the human
CC heregulin variant, gamma-HRG protein, isolated from a native HRG source,
CC such as human MDA-MB-175 cells. It has atleast one biological property
CC as the native sequence.
XX
XX Sequence 768 AA;
SQ
Query Match 39.2%; Score 790.5; DB 21; Length 768;
Best Local Similarity 41.5%; Pred. No. 4.7e-57;
Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;
QY 69 NSYLLRACSGPOQASSGPPNHHSSQSTLRPPPLPP-PHNHTLSHHH-SSANSLNRLNLTNR 126
Db 149 NSNLTLDTEHENTETDHPGGLQNHARLTTPPPPLSHAHTPNQHHAAASINSLNRGNFTPR 208
QY 127 RQSIHAPAPND--LATTP-----ESVOLQDSWVLNSNVPLETR----- 164
Db 209 SN-----PSPAPT DSHLSGEPAGGAGQEPAAHQAENWLLNSNIPLETNRNKGKQPLGLTQDN 264
QY 165 -----HFLFKTSSGTPLEFSSSSPGYPLTSGTVYTPPPRLPRNTF 205

```

      ||||| |::||| ::||| ||||| ||||| ||||| ||||| |||||
Db 265 LIEMDILGASRHDGAYSDGHELFK-PGCTSPLEFCTTSPGYPLTSTVYSPPPRPLPRSTF 323
QY 206 SRKAFKLRKPSKYCSWKCAALSATAAALLAILLAYFI-----VPWSL----- 248
      :| || |||||::||| ||||| : | ||||| : |
Db 324 ARPANLKKPSKYCNWKCAALSATVILLAYFVAMHLFGLNHLQPMEGOMYEIT 383
QY 249 -----KNSSIDSGEAEVGRRT 265
      ::| ||||| :||| :
Db 384 EDTASSNPVPTDVSLYPSGGTGLETPDRKGKTTEGKPSFFPEDSFIDSGEIDVGRAS 443
QY 266 QEVPPGVFWRQIHIISOPQELKENISLCKDALFGVYIRRGLPSPSHAQYDEMERLDCK--- 322
      |::||| ||||| : | ||||| || ||||| || ||||| |::| ||| :
Db 444 QKIPPGTFWRQVFIIDHPVHLKENVSLGKAALVGIYGRKGLPPSHTQDFVELLDGRLL 503
QY 323 --EKWSVYESPRRRSIQTLVQNEAVFYQYLDVGLWHLAFYNDGDKEMVSFNTVVL 377
      | : :||| : | :||| :||| ||||| :||| :
Db 504 TQEARSLGTPRQSRGTVPSPSHETGFIQYLDGSIWHLAFYNDGKESEVVSFLTAT 560
```

Search completed: August 26, 2003, 14:31:00
Job time : 60 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:29:28 ; Search time 19 Seconds
(without alignments)
1928.435 Million cell updates/sec

Title: US-09-584-411c-22

Perfect score: 2017

Sequence: 1 MLHAANKGRPSAAGRP.....NDGKDKEMVSPNTVLDGTI 381

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 853 | 42.3 | 2825 | 2 T14271 | Doc4 protein, stre |
| 2 | 205.5 | 10.2 | 2406 | 2 A54148 | od2 protein - fruit |
| 3 | 186.5 | 9.2 | 2515 | 2 S47008 | tenascin-like prot |
| 4 | 156.5 | 7.8 | 528 | 2 I47141 | gastric mucin (clo |
| 5 | 150 | 7.4 | 782 | 2 A61625 | tenascin-like prot |
| 6 | 141 | 7.0 | 2531 | 2 T16743 | proline-rich prote |
| 7 | 134.5 | 6.7 | 633 | 2 S62057 | eyelid - fruit fly |
| 8 | 132.5 | 6.6 | 2715 | 2 T13049 | C-terminal domain- |
| 9 | 132 | 6.5 | 1268 | 2 T31420 | hypothetical prote |
| 10 | 131 | 6.5 | 1892 | 2 T18314 | hypothetical prote |
| 11 | 129.5 | 6.4 | 731 | 2 T04455 | hypothetical prote |
| 12 | 129 | 6.4 | 383 | 2 T46707 | proteophosphoglyc |
| 13 | 129 | 6.4 | 1456 | 2 T01397 | LTR gag/pol polypr |
| 14 | 128 | 6.3 | 356 | 2 A96826 | TRK14.10 (imported |
| 15 | 127.5 | 6.3 | 554 | 2 T49833 | related to Vea pro |
| 16 | 127.5 | 6.3 | 1420 | 2 T37781 | probable cytoskele |
| 17 | 126.5 | 6.3 | 720 | 2 J01676 | AB13 protein - Ara |
| 18 | 126 | 6.2 | 620 | 2 S67333 | hydroxyproline-ric |
| 19 | 126 | 6.2 | 710 | 2 D96728 | hypothetical prote |
| 20 | 126 | 6.2 | 2232 | 2 T34434 | hypothetical prote |
| 21 | 124.5 | 6.2 | 731 | 2 B86369 | hypothetical prote |
| 22 | 123.5 | 6.1 | 317 | 2 S5516 | mucin (clone PGM-2 |
| 23 | 123.5 | 6.1 | 1097 | 2 T13033 | cyclin T - fruit f |
| 24 | 123.5 | 6.1 | 1140 | 2 D8690 | protein F4H10.3 (|
| 25 | 123 | 6.1 | 193 | 2 C86334 | hypothetical prote |
| 26 | 123 | 6.1 | 490 | 2 S52830 | HMS1 protein - yea |
| 27 | 122.5 | 6.1 | 968 | 2 S46992 | protein p130 - rat |
| 28 | 122.5 | 6.1 | 1680 | 2 T41628 | probable transcrip |
| 29 | 121.5 | 6.0 | 760 | 2 F86387 | probable Pto kinas |

RESULT 1

T14271

Doc4 protein, stress-induced - mouse

N:Alternate names: od2 protein homolog

C:Species: Mus musculus (house mouse)

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T14271

R:Wang, X.Z.; Kuroda, M.; Sok, J.; Batchvarova, N.; Kimmel, R.; Chung, P.; Zinszner, EMBO J. 17, 3619-3630, 1998

A:Title: Identification of novel stress-induced genes downstream of chop.

A:Reference number: Z17951; MUID:98315054; PMID:9649432

A:Accession: T14271

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2825 <N>

A:Cross-references: EMBL:AF059485; NID:g3170614; PID:g3170615; PIDN:AAC31807.1

C:Genetics:

A:Gene: Doc4

| | | | | | |
|-----------------------|-----|--|--------------------|-------------|--------------|
| Query Match | | 42.3% | Score 853; | DB 2: | Length 2825; |
| Best Local Similarity | | 40.1% | Pred. No. 2.6e-51; | | |
| Matches 193; | | Conservative 55; | Mismatches 111; | Indels 122; | Gaps 12; |
| QY | 12 | SAEAGRIPTTSSPSLLPSAQLPSSHNPPVS-CQPLLDSTSHQIMDTPDEEFSNPS | 70 | | |
| DB | 161 | NTETGAPL-HCSSASSTIEQSPSPPPANESQRRLLGNGVAQPTPDSDEEFVNS | 219 | | |
| QY | 71 | YLLRACSGPQQAASSGGPP-----NHQSQTLPRLPPPHNHTLSHHH-SNANSLNRSLTN | 125 | | |
| DB | 220 | FLVKSGSASLGVAANDHPSSLQNHPRLT--PPPLPHAHTPNQHAASINSLNRGNFTP | 277 | | |
| QY | 136 | RRSQIHAPAPND-----LATTPESVQLQDSVNLNSVPLETR----- | 164 | | |
| DB | 278 | RSN-----PSPAPTDHSLSGEPAGAQETPAQDNWVLSKIPVETRLGKQPFVGTWOD | 333 | | |
| QY | 165 | -----HFLKTSSTSGSPFLSSSPGYPPLTSGTVYTPPPRLPLPRT | 204 | | |
| DB | 334 | NLIENDIFSASRRDGAYSDGHFFFR-PGTSPLCTTSPGYLTSTTVYSPRPLRST | 392 | | |
| QY | 205 | FSRKAFKKKSKYCSWKAALSAATAALLAILLAYFI-----VPWSL----- | 248 | | |
| DB | 393 | FSRPAFLKPKSKYCNWKAALSAIISATVILLAYFVAMHLFLNHLNMPMEQOMQY | 452 | | |
| QY | 249 | -----KSSSDSGAEVGR | 262 | | |
| DB | 453 | EITEDTASWPVPTDVSLYPSGGTGLETPDRKKGAAEGKPSLSPEDSFIDSGEIDVGR | 512 | | |
| QY | 263 | RVTQVPPGVFWRSQIHISQPFKFNISLAKDALFGVYIRRGPPSHAQDFMERLDGK | 322 | | |
| DB | 513 | RASQIPPGTWRSQVFDHPVHLKFNVLGKALVGIYGRKGLPSPHTQLDFVLLDGR | 572 | | |
| QY | 323 | -----EKSVSVESPRRERSIQTLYQNEAVFYQYLDVGLWHLAFYNDGKDKEMVSFNTVVL | 377 | | |

ALIGNMENTS

Query Match 6.7%; Score 134.5; DB 2; Length 633;
Best Local Similarity 25.1%; Pred. No. 0.074;
Matches 55; Conservative 24; Mismatches 83; Indels 57; Gaps 8;
QY 18 PIPPTSSPSLLSAQLPSSHN-----PPVSCQ-----MPLDLSNTSHQ 56
DB 264 PSQPSQNPFPPIPEIPSTQSATNPFPFVQOQFNQPSMGIPQONRPLPQLPNRNNRP 323
QY 57 IMDTNPDEEFPNSYLLRACSGPQQASSGPPNNHHSQSTLRPLPPPHNHTLSHHSSAN 116
DB 324 VPPPPMRTTEGSGVRLPAPPPRRGAPP-----PPPHRVTS---NTLN 368
QY 117 SLNRSLNRRSIIHAPAPNDLATPE---SVQLQDSWVLSNVPLETRHFLFTSSG 173
DB 369 SAGNSILLQATGRRGAPPPPPPRASRPTPNVTMOQNPOQYNNRNP-----FGYQTSN 423
QY 174 STPLFSSSSPGYPLTSGTVTP-----PPRLPNT 204
DB 424 -----MSSPPPPPVTFNLTPOMTAATGQPAVLPQNT 457
RESULT 8
T13049
eyelid - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13049
R:Treisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.
A:Reference number: Z17592
A:Accession: T13049
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2715 <TRE>
A:Cross-references: EMBL:AF053091; NID:q2981220; PID:g2981221; PIDN:AAC06254.1
A:Genes: eld
C:Function:
A:Description: could act as a transcription factor antagonistic to the wg pathway
C:Keywords: DNA binding
Query Match 6.6%; Score 132.5; DB 2; Length 2715;
Best Local Similarity 24.3%; Pred. No. 0.06;
Matches 64; Conservative 21; Mismatches 97; Indels 79; Gaps 13;
QY 8 GRKPSAEAGRPPTSSPSLLPSAQLP-----SSHNPPEVS-----CQMLDLSNT 53
DB 413 GSSPSGSGHPLPPASPHVHPVLPQQPPPPPHVSAGGPPSPSGHAPSPSPQASPS 472
QY 54 SHQ-IMDTPNDEEFS-----PN-SYLLRACSGP----- 79
DB 473 PHQELIGQNSDSSSGAHSGMGSGPPTPNQOVNRPTPTSGSGSRMSPAVAQNH 532
QY 80 -----QQASSGPPNNHHSQSTLRPLPPH-----NHTLSHHSSANSNLSLN 125
DB 533 ISRPASNQSSSGPMQPPVAGGPPMPHPGMPGPPQQQSQSQASNS--ASSASN 590
QY 126 RRSQIHAPAP-----NDLATTPEVSQLODSWVLSNVPLETR-HFLFTSSGSTPLFS 180
DB 591 SPOQTTPPPPPNQGNHNNATPPPPQ-----GAAGGYPMPPHMHGKMGK---PQSP 643
QY 181 SSPGYPLTSGTVTP---PPR 198
DB 644 GAQGYPPQPPQYPPGNYPPR 664
RESULT 9
T31420
C-terminal domain-binding protein ra8 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
C:Accession: T31420

R:Yuryev, A.; Patturajan, M.; Litington, Y.; Joshi, R.V.; Gentile, C.; Gabara, M.; C
Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996
A:Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts
A:Reference number: Z21024; MUID:96293459; PMID:8692929
A:Accession: T31420
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1268 <YUR>
A:Cross-references: EMBL:U49055; NID:g1438529; PID:g1438530; PIDN:AAC52656.1
A:Experimental source: hippocampus
Query Match 6.5%; Score 132; DB 2; Length 1268;
Best Local Similarity 29.2%; Pred. No. 0.27;
Matches 63; Conservative 22; Mismatches 85; Indels 46; Gaps 14;
QY 11 PSACAGRPPTSSPSLLPS-AOLPSSHNPVSC-OMPLDLSNTSHQIMDTPNDEEFS 68
DB 778 PSCESTRPVIPSDIPSAAPMAQPPGASNTSGILCVQRPNVSSNS--EILGVSPAN--VS 833
QY 69 NSYLLRACSGPQQASSG-----PPNHHSQSTLRPLPP--PNHHTLSHHSSANSNLSLN 121
DB 834 NSAAINGAOPPPNNLNSGILGIQPPNVSSGSLGLVLPPLPN-----SGLVGLQPP 886
QY 122 SLNRRSQIHAPAP-APNDLATTPESVQLQDSWVLSNVP-----LETRHFLFTSSGSTP 176
DB 887 NVTNPAGLLGTQPPICQPNL--PPLTIPAQ-----RMPALPMLDIRPGLIAQAFC--P 935
QY 177 LFSSSSPGYPLTSGTVTPPPRL-----PRNTF 205
DB 936 RFPLLQPGIPPPQSG---IPPSVLDALHPPPPRGPF 968
RESULT 10
T18314
hypothetical protein L7610.4 - Leishmania major
C:Species: Leishmania major
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18314
R:Oliver, K.; Murphy, L.; Quail, M.; Lawson, D.; Harris, D.; Rajandream, M.; Ivans, A
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18876
A:Accession: T18314
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1892
A:Cross-references: EMBL:AL034356; NID:el371876; PID:el371559; PIDN:CAA22240.1
C:Genetics:
A:Note: L7610.4
Query Match 6.5%; Score 131; DB 2; Length 1892;
Best Local Similarity 20.3%; Pred. No. 0.53;
Matches 88; Conservative 57; Mismatches 169; Indels 120; Gaps 19;
QY 12 SABAGRPPTSSP-----SLLPSAQLPSSHP-PPVSCQMLDLSNTSHQIMDTPNDEEFS 67
DB 625 SATVAAPISKTDPYAGHTAPDGEPRPLHVMPMPPIQRPYAAT-----EEGA 673
QY 68 PNSYLLRACSGPQQASSGPPNNHHSQSTLRPLPPPHNHTLSHHSSANSNLSLN 127
DB 674 PR-FSVRYKVTAPQESDAPSPH-----SPPAHTLLSRGAADA-GEAAAKETR 721
QY 128 SQIHAPAPNDLATP-----PESVQLQ-----DSWVLSNVPLETR 164
DB 722 EQLSAAKEAVTAMTTAGVQSSKKPQOLQHEPDGSPNGDDVDELLEADLLIMQSRPQSSR 781
QY 165 HFLFTSSGSTPLFSSSSPCYP-----LTSGTVYT-----PPRLLP 201
DB 782 H---GAAGSSPQSIHSPQSPAVECFSDAPRSVHATGSETTTEOQQOQORQHPOLRP 838
QY 202 RNTFSRKAFKLPKPSKYCSWKCAALSAIAAALLAILLAYFIVPWSLKSSISDGEAEV 261
DB 839 -TSLGRGAVASDSRCRAAEKAEELTPPPPAQLL-----PTAMD-----DITERDVG 884

QY 262 RRVTOEVPVGRMSQIHISOPQLKFNLSLGDALFGVYIRRL-PPSHAYQDFM---- 316
Db 895 EQ-----SNVYIRGHMMNTLPQKMDAPRCGGSTAVHGA--TEGLEPPQDVAVDVYMGTA 936
QY 317 ERLDGKE-----KWSVVEPRRRTSTQTLVQNEAVFVQYLDVGLWHLA 359
Db 937 ELREGKEALCGPSTGADVGPKTLQEAADTSEPTGASAPLVLRERYVYVDPLEASMDAA 996
QY 360 FYNQKQKEMVSN 373
Db 997 YMLDTLEQQYVRIN 1010

RESULT 11

T04455
Hypothetical protein F4D11.90 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 04-Mar-2000
C:Accession: T04455
R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Hohelsel, J.; Mewes, H.W.
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15360
A:Accession: T04455
A:Molecule type: DNA
A:Residues: 1-731 <BEV>
A:CROSS-references: EMBL:AL022537
A:Experimental source: cultivar Columbia; BAC clone F4D11
C:Genetics:
A:Map position: 4
A:Introns: 326/1; 476/1; 505/1; 528/3; 557/2; 606/3; 660/3
A:Note: F4D11.90
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match 6.4%; Score 129.5; DB 2; Length 731;
Best Local Similarity 23.5%; Pred. No. 0.2;
Matches 93; Conservative 33; Mismatches 148; Indels 121; Gaps 17;

QY 11 PSAEAGRIPIPTSSPSLLPSAQLPSS-HNPPVSCOMPLDLSNTSHQIMDT-----NPDEE 65
Db 21 PPADS---VPDTSSPPAPPLSPPLPPLSPPLSPPLSPPLSPPLSPPLSPPLSPPLSPPLSPPTIE 77
QY 66 FSPNSVLLRACSGPQOASSGPPNHHQSOTLRPLP-----PPHNHTLSHHSSANSL 118
Db 78 -SPPPLLESPPPLESP 136
QY 119 N-----RNSLTNRSSQIHAPAPA-----PNDLATTTPESVOL 149
Db 137 SPPRSLPSESTPPVNTASPPPPPPRRS---GPKPSPPPPINSPPNPSNTFSLPET 193
QY 150 QDSWVLSNVPLETRHFLFKTSSTGTPLESS-----SPGYPLSGTYVTP----- 195
Db 194 SPP-----PKPPLSTTTPF-----PSSSTPPPKSPAAVTLPPFGPAGQLPDGTVAPPIGRVI 245
QY 196 -----PPRLPRNTERKAFKLPKSKYCSWKCAAL--SAIAAALLAIL 239
Db 246 EPKTSPASISPGTQPLVPKS-----LPVTSYHRSSAGFLGGVIGVALLILG 297
QY 240 AYFIVPWSLKNSSIDSGAEVGRRTQVEP---PGVWRSSQIH-ISOQFLKFNISLGD 295
Db 298 LLEFVYRATRNRNNNSAAHQSKTPSKVQHRGNGAGTNAHVITMP----- 345
QY 296 ALFGVYIRGLPPSHQAQDFMERLDGKEKWSVES 330
Db 346 -----PPIHAKYISGGCDTKENNSVAKN 369

RESULT 12

T46707
proteophosphoglycan, membrane-associated [imported] - Leishmania major (fragment)
C:Species: Leishmania major
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 11-May-2000
C:Accession: T46707
R:Ilg, T.; Montgomery, J.; Stierhof, Y.D.; Handman, E.

J. Biol. Chem. 274, 31410-31420, 1999
A:Title: Molecular cloning and characterization of a novel repeat-containing Leishman
a1dylinositol anchor.
A:Reference number: Z23133; MUID:20002668; PMID:10531342
A:Accession: T46707
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-383 <ILG>
A:CROSS-references: EMBL:AJ243460; PIDN:CAB46680.1
A:Experimental source: strain LRC-1137
C:Genetics:
A:Gene: pp91
C:Function:
A:Description: may be important for direct host-parasite interactions

Query Match 6.4%; Score 129; DB 2; Length 383;
Best Local Similarity 25.0%; Pred. No. 0.095;
Matches 51; Conservative 34; Mismatches 91; Indels 28; Gaps 4;
QY 4 AANKGRKPSAEAGRIPIPTSS-PSLLPSAQLPSSHNPPVSCOMPLDLSNTSHQIMDTNP 62
Db 74 SASSSAPSSSSSSAPSSASSSAPSSSSAPSSASSSAPSSSSAPSSASSSASSS 124
QY 63 DEEFSPNSVLLRACSGPQOASSGPPNHHQSOTLRPLP-----PPHNHTLSHHSSANSLNRNS 122
Db 125 ----APSS-----SSAPSSASSSAPSSSSAPSSASSSAPSSSSAPSSASSSAPSS 176
QY 123 LTNRRSQIHAPAPAPNDLATTPEVSQLOQDSWVLSNVPLETRHFLFKTSSTGTPLEFSSSS 182
Db 177 SSSAPSSASSSAPSSSSSSAPSSAS 250
QY 183 PGYPLTSGTYVTPPPRLPLRNTFS 206
Db 227 PSSSTTTTMDPTDPVLPSSSSS 250

RESULT 13

T01397
LIR gag/pol polyprotein homolog T419.16 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 24-Nov-1999
C:Accession: T01397
R:Parnell, L.D.; Gnoj, L.; de la Bastide, M.; Hameed, A.; Habermann, K.; Schutz, K.;
submitted to the EMBL Data Library, May 1998
A:Description: Genomic sequence of BAC T419 from Arabidopsis thaliana, chromosome IV,
A:Reference number: Z14314
A:Accession: T01397
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1456 <PAR>
A:CROSS-references: EMBL:AF069442; NID:G3242970; PID:G3924609
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 129/1
A:Note: T419.16
C:Superfamily: retrovirus-related polyprotein

Query Match 6.4%; Score 129; DB 2; Length 1456;
Best Local Similarity 22.0%; Pred. No. 0.52;
Matches 87; Conservative 42; Mismatches 147; Indels 120; Gaps 17;
QY 11 PSAEAGRIPIPTSSPSLLPSAQLPSSHNPPVSCOMPLDLSNTSHQIMDTPDEEFPNS 70
Db 758 PTTPLVLPAPPCLGPHLDTSPRPPSPSP---LCITQVSSSNL-----PSS 800
QY 71 YLLRACSGPQOASSGPPNHHQSOTLRPLP-----PPHNHTLSHHSSANSLNRNSLTNRROI 130
Db 801 ----SISPSSEPTAPSHNGPQTAQP-----HOTQNS-NSNSPILNPNP 842
QY 131 HAPAP-APNDLATTPEVSQLOQDSWVLSNVPLETRHFLFKTSSTGTPLEFSSSPGYPLTS 189
Db 843 NSPSPNSPNQNSPLPQS-----PISSPHI-----PTPSTSISEPNSSPS 882

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:28:28 ; Search time 15 seconds
(without alignments)
1194.478 Million cell updates/sec

Title: US-09-584-411c-22

Perfect score: 2017

Sequence: 1 MHAANKGRKPSAAGRP.....NDGRDKEMSVNTVLDGTI 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------|--------------------|
| 1 | 134.5 | 6.7 | 633 | 1 LA17_YEAST | Q12446 saccharomyc |
| 2 | 134 | 6.6 | 1822 | 1 ZAP3_HUMAN | P49750 homo sapien |
| 3 | 126.5 | 6.3 | 720 | 1 ABI3_ARATH | Q01593 arabidopsis |
| 4 | 126 | 6.2 | 620 | 1 EXTN_TOBAC | P13983 nicotiana t |
| 5 | 125 | 6.2 | 852 | 1 WS14_HUMAN | Q9np71 homo sapien |
| 6 | 124.5 | 6.2 | 827 | 1 M4K1_MOUSE | P70218 mus musculu |
| 7 | 123.5 | 6.1 | 1097 | 1 CCT_DROME | O96433 drosophila |
| 8 | 123 | 6.1 | 490 | 1 MOT3_YEAST | P54785 saccharomyc |
| 9 | 123 | 6.1 | 1157 | 1 BIC1_YEAST | P47068 saccharomyc |
| 10 | 121 | 6.0 | 1509 | 1 GSR1_HUMAN | Q9nzm4 homo sapien |
| 11 | 121 | 6.0 | 5179 | 1 MUC2_HUMAN | Q02817 homo sapien |
| 12 | 120.5 | 6.0 | 701 | 1 CG1_HUMAN | Q13495 homo sapien |
| 13 | 120.5 | 6.0 | 968 | 1 BCAL_RAT | Q63767 rattus norv |
| 14 | 119.5 | 5.9 | 649 | 1 V7OK_EPMV | P20129 eggplant mo |
| 15 | 117.5 | 5.8 | 894 | 1 M4K3_HUMAN | Q8ivh8 homo sapien |
| 16 | 117.5 | 5.8 | 1367 | 1 AMTH_YEAST | P08640 saccharomyc |
| 17 | 117 | 5.8 | 860 | 1 ARCA_PENRO | Q13508 penicillium |
| 18 | 117 | 5.8 | 2471 | 1 NTC2_HUMAN | Q04721 homo sapien |
| 19 | 116.5 | 5.8 | 2556 | 1 NTC1_HUMAN | P46531 homo sapien |
| 20 | 116 | 5.8 | 788 | 1 PCAP_HUMAN | Q96rn5 homo sapien |
| 21 | 115.5 | 5.7 | 1003 | 1 MED6_HUMAN | Q96dn6 homo sapien |
| 22 | 115 | 5.7 | 684 | 1 RPI1_HUMAN | P78424 homo sapien |
| 23 | 115 | 5.7 | 833 | 1 M4K1_MOUSE | Q92918 homo sapien |
| 24 | 115 | 5.7 | 2471 | 1 NTC2_RAT | Q9qwt3 rattus norv |
| 25 | 114.5 | 5.7 | 413 | 1 ESC1_SCHPO | Q04635 schizosacch |
| 26 | 114.5 | 5.7 | 5085 | 1 PCLO_RAT | Q91ks6 rattus norv |
| 27 | 114 | 5.7 | 1006 | 1 P5PA_HUMAN | Q15735 homo sapien |
| 28 | 114 | 5.7 | 3421 | 1 TEGU_HSVB | P28955 equine herp |
| 29 | 113.5 | 5.6 | 513 | 1 M113_MOUSE | Q8bgt6 mus musculu |
| 30 | 113.5 | 5.6 | 676 | 1 MUC1_MESAU | Q60528 mesocricetu |
| 31 | 113.5 | 5.6 | 1132 | 1 YK5_YEAST | P34250 saccharomyc |
| 32 | 113.5 | 5.6 | 1183 | 1 DRPL_RAT | P54258 rattus norv |
| 33 | 113 | 5.6 | 890 | 1 HCNI1_HUMAN | O60741 homo sapien |

RESULT 1

| ID | LA17_YEAST | STANDARD; | PRT; | 633 AA. |
|----|---|-----------|------|-----------|
| AC | Q12446; | | | |
| DT | 01-NOV-1997 (Rel. 35, Created) | | | |
| DT | 01-NOV-1997 (Rel. 35, Last sequence update) | | | |
| DT | 01-NOV-1997 (Rel. 35, Last annotation update) | | | |
| DE | Proline-rich protein LAS17. | | | |
| GN | LAS17 OR YOR181W. | | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomycetes. | | | |
| OX | NCBI_TaxID=4932; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | Toh-E A.; | | | |
| RL | Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases. | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | Hughes B., Pohl T.M.; | | | |
| RL | Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases. | | | |
| CC | -!- SIMILARITY: TO S.POMBE WSP1. | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | | |
| CC | EMBL; D78487; BAAL1386.1; -. | | | |
| DR | EMBL; 275089; CAA99390.1; -. | | | |
| DR | PIR; S62057; S62057. | | | |
| DR | SGD; S0005707; LAS17. | | | |
| DR | GO; GO:0005857; C:actin cortical patch (sensu Saccharomycetes); IDA. | | | |
| DR | GO; GO:0005737; C:cytoplasm; IDA. | | | |
| DR | GO; GO:0008092; P:cytoskeletal protein binding activity; IDA. | | | |
| DR | GO; GO:0007015; P:actin filament organization; IDA. | | | |
| DR | GO; GO:0008154; P:actin polymerization and/or depolymerization; IDA. | | | |
| DR | GO; GO:0016288; P:cytokinesis; IMP. | | | |
| DR | GO; GO:0006897; P:endocytosis; IMP. | | | |
| DR | GO; GO:0007121; P:polar budding; IMP. | | | |
| DR | GO; GO:0008970; P:response to osmotic stress; IMP. | | | |
| DR | InterPro; IPR000697; EVH1. | | | |
| DR | InterPro; IPR001960; WH1. | | | |
| DR | InterPro; IPR003124; WH2. | | | |
| DR | Pfam; PF005568; WH1; 1. | | | |
| DR | Pfam; PF02205; WH2; 1. | | | |
| DR | SMART; SM00461; WH1; 1. | | | |
| DR | SMART; SM00246; WH2; 1. | | | |
| FT | DOMAIN 185 190 | | | POLY-PRO. |
| FT | DOMAIN 323 329 | | | POLY-PRO. |
| FT | DOMAIN 342 348 | | | POLY-PRO. |
| FT | DOMAIN 352 358 | | | POLY-PRO. |
| FT | DOMAIN 385 391 | | | POLY-PRO. |

O35449 mus musculu
P37370 saccharomyc
Q92412 rattus norv
P46938 mus musculu
Q924h2 mus musculu
Q9wub5 rattus norv
Q8wxx7 homo sapien
P10358 turnip yell
P28477 turnip yell
P06599 daucus caro
P06601 drosophila
Q99mz3 mus musculu

ALIGNMENTS

FT DOMAIN 427 431 POLY-PRO.
 FT DOMAIN 470 474 POLY-PRO.
 FT DOMAIN 503 507 POLY-PRO.
 FT DOMAIN 520 526 POLY-PRO.
 SQ SEQUENCE 633 AA; 67571 MW; 4488355563AA2645 CRC64;

Query Match 6.7%; Score 134.5; DB 1; Length 633;
 Best Local Similarity 25.1%; Pred. No. 0.051;
 Matches 55; Conservative 24; Mismatches 83; Indels 57; Gaps 8;

QY 18 PIPPTSPSLPSAQLPSSHN-----PPVSCO-----MPLDLSNTSHQ 56
 DB 264 PSQSQSNPPFPPEIPSTQSATNPPFPVPOQFNOAPSNGIPQQRNPQLPNNRP 323
 QY 57 IMDTPNDEEFPSPNYSLLRACSGQQASSGPPNNHSSQSTLRPLPPPHNHTLSHHSSAN 116
 DB 324 VPPPPMRTTTEGSGVRLPAPPPPPRRGPAPP-----PPPHRHVTS---NTLN 368
 QY 117 SLNRNLSITNRSOIHAPAPAPNDLATPE---SVQLQDSWVLSNVPLETRHFLFKTSSG 173
 DB 369 SAGGNSLLPQATGRRGPAPPPPPRRASRPTPNVTMOQNPQYNNRNP-----FGYQTN 423
 QY 174 STPLFSSSSPGYPLTSGTVTP-----PPRLPNT 204
 DB 424 -----MSSPPPPPTVTNTLTPQNTAATGQAPVLPQNT 457

RESULT 2

ZAP3_HUMAN STANDARD; PRT; 1822 AA.
 ID ZAP3_HUMAN
 AC P49750; P49752; Q9P1V7;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE "Sequencing of human chromosome 14q24.3 region."
 GN Nuclear protein ZAP3 (ZAP113).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rowen L., Madan A., Qin S., Abbasi N., Baradarani L., Birditt B.,
 RA Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,
 RA Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.,
 RA "Sequencing of human chromosome 14q24.3 region."
 RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.
 RP TISSUE=Brain;
 RC MEDLINE=95319502; PubMed=7596406;
 RX Sherrington R., Rogaeve E.I., Liang Y., Rogaeve E.A., Levesque G.,
 RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
 RA Poncin J.-F., Brunl A.C., Montesi M.P., Sorbi S., Rainero I.,
 RA Pinesse L., Nee L., Chumakov I., Pollen D., Brookes A.,
 RA Sansau P., Polinsky R.J., Wasco W., da Silva H.A.R., Haines J.L.,
 RA Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
 RA Rommens J.M., St George-Hyslop P.H.;
 RT "Cloning of a gene bearing missense mutations in early-onset familial
 RT Alzheimer's disease."
 RL Nature 375:754-760(1995).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
 CC frameshift in position 1661.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AC007956; AAF61275.1; -
 DR EMBL; L40403; AAC42008.1; ALT_FRAME.
 DR EMBL; L40400; AAC42006.1; -
 KW Nuclear protein.
 FT DOMAIN 15 205 PRO-RICH.
 FT DOMAIN 382 430 GLN-RICH.
 FT DOMAIN 807 1209 ARG-RICH.
 FT DOMAIN 1488 1577 ARG-RICH.
 FT CONFLICT 621 621 P -> S (IN REF. 2).
 FT CONFLICT 1404 1404 T -> I (IN REF. 2).
 FT CONFLICT 1821 1821 K -> E (IN REF. 2).
 SQ SEQUENCE 1822 AA; 204947 MW; 8E6CB83FE540C7D2 CRC64;

Query Match 6.6%; Score 134; DB 1; Length 1822;
 Best Local Similarity 26.6%; Pred. No. 0.2;
 Matches 53; Conservative 22; Mismatches 92; Indels 32; Gaps 7;

QY 20 PPTSSSLLPSAQLPSSHNPPVPS--COMPLDLSNTSHQIMDTPNDEEFPSPNYSLLRACS 77
 DB 176 PPTSSQPYLPAPQSPSPSQSYLAPTSYSSSSSSQSYLSHSQSYLPSS----- 228
 QY 78 GPOQASSGPPNNHSSOSTL---RPPLPPPHNHTLSHHSSANSLSNRRSLNRRSOIHAPA 134
 DB 229 ---QASPSRFSQGHKSQQLLAPPSPAPPNGKTVQOEPL-----ESGANKSTEQQA 280
 QY 135 PAPNDLATTPESVQLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSP-----GYPL 187
 DB 281 PEPDPSTMTQ---EQQYWYRQHLHLSLQRTKVLPHGKGVVAKDTPPEVKEEVTVA 338
 QY 188 TSGTVTP---PPRLPNT 203
 DB 339 TSQVPSPSPSEEPPLPPN 357

RESULT 3

ABI3_ARATH STANDARD; PRT; 720 AA.
 ID ABI3_ARATH
 AC Q01593;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Abscissic acid-insensitive protein 3.
 GN ABI3 OR A13G24650 OR MSD24.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Fruit;
 RX MEDLINE=93076109; PubMed=1359917;
 RA Giraudat J., Hauge B.M., Valon C., Smalle J., Parcy F.,
 RA Goodman H.M.;
 RT "Isolation of the Arabidopsis ABI3 gene by positional cloning."
 RL Plant Cell 4:1251-1261(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20363039; PubMed=10907853;
 RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
 RT and BAC clones."
 RL DNA Res. 7:217-221(2000).
 CC -1- FUNCTION: COULD PARTICIPATE IN ABSICISIC ACID-REGULATED GENE
 CC EXPRESSION DURING SEED DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: Contains 1 TF-B3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X68141; CAA48241.1; -.
 DR EMBL: AF000740; BAB01214.1; -.
 DR PIR: JQ1676; JQ1676.
 DR InterPro: IPR003340; TF_B3.
 DR Pfam: PF02362; B3; 1.
 KW Developmental protein; Transcription regulation; DNA-binding;
 KW Activator; Nuclear protein. TF-B3.
 FT DOMAIN 532 720
 SQ SEQUENCE 720 AA; 79500 MW; D6A0F740D733060F CRC64;

Query Match 6.3%; Score 126.5; DB 1; Length 720;
 Best Local Similarity 21.6%; Pred. No. 0.21;
 Matches 77; Conservative 40; Mismatches 129; Indels 111; Gaps 17;
 QY 20 PPTSSPLLSAQLPSSNNPPVSCMPLLDSTSHQIMDNPDEEF----- 66
 Db 378 PP-QPEFLPLESPSPPPPPSGMP-----HQOFPMPTSQYNQFGDPTGFNGYN 428
 QY 67 -SPNSY-----LLRACSGPQQAASSGPPNHHQSQTLRPLPPPHNHTLSHHHS 113
 Db 429 MNPYQYVPVAGQMRDQRLRLCSATKEA-----RKRMARQRRLSHHHR 475
 QY 114 SANSLNRLNRLNRRSIQH-----APAPAPNDLATTPEVQLQDSWNLNSNVPTRHFL 167
 Db 476 HNNN-NNNNNNQOQTGETCAVAQPLNPVATATG-----GTWTFWNPVAPPQL- 529
 QY 168 FKTSSGTPLFSSSPGYPVLSGVYTPPRLPNTFSKAFKLPKSKYCSWKCAALS 227
 Db 530 -----PPVMTQLPTMDRAGSASAPRQOVVPR---RQGWKPEKRLFLQLVKQS 579
 QY 228 ATAAALLAILLAYIVPWSLKNSSIDSGEAEVGRVTVQVPPGV-----FW--- 275
 Db 580 DVGN-----LGRVLPRKAEATHLEARDGSLAME-DIGTSRVNMYRFRWPNNK 631
 QY 276 SQIHI--SOPOKLFN-----ISLGDALFGVYIRRL-----PPSHA 311
 Db 632 SRMYLLENTGDFVKTNGLEGDFVIYSDVKCGKVLIRGVKVPQSGQKPEAPPSSA 688

RESULT 4
 EXTN_TOBAC
 ID EXTN_TOBAC STANDARD; PRT; 620 AA.
 AC P13983;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
 GN HRGPNT3.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Xanthi; TISSUE=Leaf;
 RX MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich
 glycoprotein gene in lateral root initiation.";
 RL Genes Dev. 3:1639-1646(1989).
 CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 MAIN ROOT.
 CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 GLYCOSYLATED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X13885; CAA32090.1; -.
 DR PIR: S06733; S06733.
 DR InterPro: IPR000480; Glutelin.
 DR PRINTS: PR002965; P_Rich_extensn.
 DR PRINTS: PR00211; GLUTELIN.
 DR PRINTS: PR01217; PRICHEXTENS.
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
 KW Hydroxylation. 1 20
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 620 EXTENSIN.
 FT REPEAT 70 73 H-A-P-P.
 FT REPEAT 148 151 H-A-P-P.
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 229 235 1.
 FT REPEAT 236 242 2.
 FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
 FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
 SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;

Query Match 6.2%; Score 126; DB 1; Length 620;
 Best Local Similarity 24.1%; Pred. No. 0.19; Mismatches 17; Gaps 11;
 Matches 59; Conservative 17;
 QY 3 HAANKGRKPSAAGRPIPTSPSLPSAQLPSSHN-----PPVSCMPLLD 51
 Db 133 HLPSSGQPPSPSHGAPP-SGHTPPRQGHPPSHRRPSPSRHGHGPPPTTAQP 187
 QY 52 NTSHOIMDTNDEEFSNYSLLRACGPO---QASSGPPNHHSTLRP-----PLPP 102
 Db 188 -----PTPIYSPS-----PQVQPPPTYSPPPTTHVQTPSPSRGHQP 228
 QY 103 PHNHT-LSHHSSANSNLSNLTNRSSQIHAPAPNDLATTPEVQLQDSWNLNSNVL 161
 Db 229 THRHPAPTHRHA-----PTHQPSRLHLPPSRQPPPTSPSPPP 270
 QY 162 ETRHFLKTSSTGSLFSSSPGY-PLTSGTVYTPPRL-----LPRNTFSRKAFLK 213
 Db 271 -----AYAQSPQSPSTYSPPPTSPSPPIYSPPPPPSPPTPTPTSPSPPPAYS 325
 QY 214 KPSKY 218
 Db 326 PPPTY 330

RESULT 5
 WS14_HUMAN
 ID WS14_HUMAN STANDARD; PRT; 852 AA.
 AC Q9NF71; Q96E48; Q9BY03; Q9BY04; Q9BY05; Q9BY06; Q9Y2P3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Williams-Beuren syndrome chromosome region 14 protein (WS basic-helix-
 loop-helix leucine zipper protein) (WS-BHLH) (Mlx interactor).
 GN WBSR14 OR MIO.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20241700; PubMed=10780788;
 RA de Luis O., Valero M.C., Perez Jurado L.A.;
 RT "WBSR14, a putative transcription factor gene deleted in Williams-
 Beuren syndrome: complete characterisation of the human gene and the

RT mouse ortholog.;

RL Eur. J. Hum. Genet. 8:215-222(2000).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND SUBUNIT.

RX PubMed=11230181;

RA Cairo S., Merla G., Urbinati F., Ballabio A., Reymond A.;

RT WBSR14, a gene mapping to the Williams-Beuren syndrome deleted

RT region, is a new member of the Mlx transcription factor network.;

RL Hum. Mol. Genet. 10:617-627(2001).

RN [3]

RP SEQUENCE OF 620-852 FROM N.A. (ISOFORM 4).

RX MEDLINE=99075645; PubMed=9860302;

RA Meng X., Lu X., Li Z., Green E.D., Massa H., Trask B.J., Morris C.A.;

RT Keating M.T.;

RT Complete physical map of the common deletion region in Williams

RT syndrome and identification and characterization of three novel

RT genes.;

RL Hum. Genet. 103:590-599(1998).

RN [4]

RP SEQUENCE FROM N.A. (ISOFORM 6).

RC TISSUE=Eye;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.;

RA Klausner R.D., Collins F.S., Wagner L., Shenman C.M., Schuler G.D.;

RA Altschul S.F., Zengberg B., Buetow K.H., Schaefer C.F., Bhat N.K.;

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.;

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.;

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.;

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.;

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.;

RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.;

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.;

RA Viallon D.K., Munzy D.M., Sodergren E.J., Lu X., Gibbs R.A.;

RA Raheij J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.;

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.;

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.;

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.;

RA Butterfield J.S., Krzywinski M.I., Skalska U., Smallus D.E.;

RA Schercher A., Schein J.E., Jones S.J.M., Marra M.A.;

RT Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.;

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -1- FUNCTION: Transcriptional repressor. Binds to the canonical and

CC non-canonical E box sequences 5'-CACGG-3' (by similarity).

CC -1- SUBUNIT: Binds DNA as a heterodimer with TCFL4/MLX.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- ALTERNATIVE PRODUCTS:

CC Event-Alternative splicing; Named isoforms=6;

CC Name=1; Synonyms=Alpha;

CC IsoId=Q9NP71-1; Sequence=Displayed;

CC Name=2; Synonyms=Beta;

CC IsoId=Q9NP71-2; Sequence=VSP_002171;

CC Name=3; Synonyms=Gamma;

CC IsoId=Q9NP71-3; Sequence=VSP_002170;

CC Name=4; Synonyms=Delta;

CC IsoId=Q9NP71-4; Sequence=VSP_002170, VSP_002171;

CC Name=5; Synonyms=Epsilon;

CC IsoId=Q9NP71-5; Sequence=VSP_002168, VSP_002169;

CC Name=6;

CC IsoId=Q9NP71-6; Sequence=VSP_002167, VSP_002172, VSP_002173;

CC TISSUE SPECIFICITY: EXPRESSED IN LIVER, HEART, KIDNEY, CEREBELLUM

CC AND INTESTINAL TISSUES.

CC -1- DISEASE: Haploinsufficiency of WBSR14 may be the cause of certain

CC cardiovascular and musculo-skeletal abnormalities observed in

CC Williams-Beuren syndrome (WBS) [MIM:194050], a rare developmental

CC disorder. It is a contiguous gene deletion syndrome involving

CC genes from chromosome band 7q11.23.

CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF

CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF156673; AAF68176.1; -

DR EMBL; AF156603; AAF68174.1; -

DR EMBL; AF245470; AAK20935.1; -

DR EMBL; AF245471; AAK20936.1; -

DR EMBL; AF245472; AAK20937.1; -

DR EMBL; AF245473; AAK20938.1; -

DR EMBL; AF245474; AAK20939.1; -

DR EMBL; AF056184; AAD28084.1; -

DR EMBL; BC012925; AAH12925.1; -

DR HSP; P25912; LHLO.

DR Genew; HGNC:12744; WBSR14.

DR MTM; 605678; -

DR MIM; 194050; -

DR GO; GO:0005667; C:transcription factor complex; NAS.

DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.

DR GO; GO:0006350; P:transcription; NAS.

DR InterPro; IPR001092; HLH_basic.

DR Pfam; PF00010; HLH; 1.

DR SMART; SM00353; HLH; 1.

DR PROSITE; PS00038; HLH_1; 1.

DR PROSITE; PS00888; HLH_2; 1.

KW Transcription regulation; Repressor; Nuclear protein; DNA-binding;

KW Williams-Beuren syndrome; Alternative splicing.

FT DOMAIN 387 394 POLY-PRO.

FT DOMAIN 409 417 POLY-PRO.

FT DNA_BIND 648 662 BASIC DOMAIN.

FT DOMAIN 688 702 HELIX-LOOP-HELIX MOTIF (POTENTIAL).

FT DOMAIN 703 724 LEUCINE-ZIPPER.

FT VARSP_LIC 176 268 Missing (in isoform 6).

FT VARSP_LIC 558 575 /FTid=VSP_002167

FT VARSP_LIC 576 852 QETVPEPCTFLPPTAP -> AVNGCQGTSPCQALGL

FT VARSP_LIC 576 852 (in isoform 5).

FT VARSP_LIC 647 648 Missing (in isoform 5).

FT VARSP_LIC 687 705 /FTid=VSP_002169.

FT VARSP_LIC 814 815 Missing (in isoform 3 and isoform 4).

FT VARSP_LIC 816 852 /FTid=VSP_002170.

FT VARSP_LIC 816 852 Missing (in isoform 2 and isoform 4).

FT VARSP_LIC 816 852 TV -> ST (in isoform 6).

FT VARSP_LIC 816 852 /FTid=VSP_002171.

FT VARSP_LIC 816 852 Missing (in isoform 6).

FT VARSP_LIC 816 852 /FTid=VSP_002172.

FT VARSP_LIC 816 852 Missing (in isoform 6).

FT VARSP_LIC 816 852 /FTid=VSP_002173.

FT CONFLICT 558 558 MISSING (IN REF. 4).

FT CONFLICT 852 852 MISSING (IN REF. 4).

SQ SEQUENCE 852 AA; 93071 MW; D49E5C3D7C0A72EC CRC64;

Query Match 6.2%; Score 125; DB 1; Length 852;

Best Local Similarity 22.2%; Pred. No. 0.32;

Matches 68; Conservative 23; Mismatches 89; Indels 126; Gaps 12;

QY 11 PSAAGRPIPTTSSPLLPSAQLPSSHNPPVSCOMPLDLSNTSHQIMDNPDEEFP-- 68

DB 498 PSRPGKASPTTAPATASPTTAGSNP---CLTQLTFA-----AKPQALEPPL 545

QY 69 -NSYLLRACSGPQOASSSGPPNHSSTL-RPPLPPPHNITLSHHSSANSLSNLTNR 126

DB 546 VSTLLRSPGSPQETVPEPCTFLPPTAPTPPPPPGATLAP-----SRPLLPVK 597

QY 127 RSGIHAPAPNDLATTPESVOLQDSWVLSNVPLETRHFLFKTSSTPLFS---SSSP 183

DB 598 AERLSPAP-----SGSERLSGDLSSMP 621

QY 184 GYPLTSGTVYTPPRLPRLPNTFSRKAFKKPKSKYCKWCAALSAIAALLAILLAYFI 243

DB 622 G-PGTLVSRVSPQPLSRG-----RP----- 642

QY 244 VPMSLNSISDSEAEVGRVTOEVPVPGVFWRSQIHISQPOFLKFNISLGKDALFGVYIR 303

ID AC Q96433; STANDARD; PRT: 1097 AA.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cyclin T.
 GN Cyt.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 KY MEDLINE=98256315; PubMed=9593731;
 RP Peng J.-M., Marshall N.F., Price D.H.;
 RT Identification of a cyclin subunit required for the function of
 RT Drosophila P-TEFb.;
 RL J. Biol. Chem. 273:13855-13860(1998).
 CC -!- FUNCTION: REGULATORY SUBUNIT OF THE CYCLIN-DEPENDENT KINASE PAIR
 CC (CDK9/CYCLIN T) COMPLEX, ALSO CALLED POSITIVE TRANSCRIPTION
 CC ELONGATION FACTOR B (P-TEFb), WHICH IS PROPOSED TO FACILITATE THE
 CC TRANSITION FROM ABORTIVE TO PRODUCTION ELONGATION BY
 CC PHOSPHORYLATING THE CTD (CARBOXY-TERMINAL DOMAIN) OF THE LARGE
 CC SUBUNIT OF RNA POLYMERASE II (RNAP II).
 CC -!- SUBUNIT: ASSOCIATES WITH CDK9 TO FORM P-TEFb.
 CC -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN C SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AF051933; AAC73052.1; -
 CC PIR; T13033; T13033.
 CC FlyBase: FBgn0025455; Cyt.
 CC GO; GO:0005634; C:nucleus; IDA.
 CC GO; GO:0005703; C:polytene chromosome puff; IDA.
 CC GO; GO:0008024; C:transcription elongation factor complex b; IPI.
 CC GO; GO:0003711; F:transcriptional elongation regulator activity; IDA.
 CC GO; GO:0009408; P:response to heat; IDA.
 CC GO; GO:0006350; P:transcription; IDA.
 CC InterPro: IPR006670; Cyclin.
 CC InterPro: IPR006671; Cyclin_N.
 CC Pfam: PF00134; cyclin; 1.
 CC SMART; SM00385; CYCLIN; 2.
 CC Cyclin; Transcription regulation.
 CC DOMAIN 15 21 POLY-SER.
 CC FT 382 390 POLY-SER.
 CC FT 426 429 POLY-SER.
 CC FT 441 449 POLY-SER.
 CC FT 553 556 POLY-PRO.
 CC FT 598 603 POLY-GLN.
 CC FT 872 880 POLY-LYS.
 CC FT 993 998 POLY-GLY.
 CC SEQUENCE 1097 AA; 118401 MW; AD55F3F57BCD3D6B CRC64;
 Query Match 6.1%; Score 123.5; DB 1; Length 1097;
 Best Local Similarity 20.5%; Pred. No. 0.55;
 Matches 78; Conservative 46; Mismatches 131; Indels 125; Gaps 15;
 QY 9 RKPSAEGRPPTPPSSPLPSAQLPSSHPPPVSCVCPPLDNTSHQINDTNPDEPSP 68
 DB 371 QRSSVSGVPGSSSSSS-----SSHKMPNYPGMP-----PEAHPDHKSQ 412
 QY 69 NSYLLRACSGPQQAASSG-----PPNHQSSTLRPPL- 100
 DB 413 PCYNNRPSSHQSSSGSLGSGSGSHSSSSSSSQQPCRFMSMPVDYHKSRGMPVVG 472
 QY 101 ---PPPHNTHSHHSSANSLNRLNLTNRSSQIHAPAPNDLATTTPESVQLQDSWYLN 158

```

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW Nuclear protein; Zinc_finger; Metal-binding; Repeat.
FT ZN_FING 346 368 C2H2-TYPE 1.
FT ZN_FING 374 397 C2H2-TYPE 2.
FT DOMAIN 8 35 POLY-GLN.
FT DOMAIN 98 104 POLY-ASN.
FT DOMAIN 143 157 POLY-ASN.
FT DOMAIN 173 177 POLY-ALA.
FT DOMAIN 240 245 POLY-HIS.
FT DOMAIN 417 420 POLY-SER.
FT DOMAIN 421 433 POLY-ASN.
FT DOMAIN 441 450 POLY-ALA.
SQ SEQUENCE 490 AA; 54382 MW; 40D0A8DE43F171ED CRC64;

Query Match
Best Local Similarity 6.1%; Score 123; DB 1; Length 490;
Matches 62; Conservative 38; Mismatches 95; Indels 54; Gaps 15;

QY 22 TSSPSLLPSAQLPSHNPVPCQPLLDSTNSHQIMTDNDEEFPNSYL-----LRA 75
DQ 164 TAAANMNAANAAAYSPFTAMPIPOD-----QQYM-----FNPASYISHYSAYNS 212
QY 76 CSGPOAASSGPPNHHSSOTLRPLPPPHNTLSHHSSA-NSLNRSILTNRRSIIHAPA 134
DQ 213 NNGCNNAANG-SNNSHSAAPAPCPPHH---HHNSTHNLNNGCAVNTN---NAPQ 265
QY 135 PAPNDLATTPEVQLQD---SWVLSNV-PLETRH---FLFKTSSGSTPLFSSSSPGYPL 187
DQ 266 HHPT-IITDQFOFQIQNPSPNLNINPAQLPLPGCKINTMPQPRTPAPNHPAPV 324
QY 188 TS-----GVVYTPP-----RLPRNTFSKAKLKPSPYCSWKCAALSAIAALLAILL 239
DQ 325 PSSNPVSNLVPAPSDHKYIHQCFCEKFKRK-----SW-----LKRHLLSHQ 371
QY 240 AVFIVPWSL 248
DQ 372 RHFLCPWCL 380

RESULT 9
ID BBCL_YEAST STANDARD; PRT: 1157 AA.
AC P47068; P47067; Q8X1F4;
DT 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin tail region-interacting protein Wt1l (Bbcl1 protein).
GN Bbcl1 OR Wt1l OR YJL020C/YJL021C OR J1305/J1286.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA To Van D., Perca J., Jacq C.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 699-765 FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR
RP LOCATION.
RX MEDLINE=21898311; PubMed=11901111;
RA Mochida J., Yamamoto T., Fujimura-Kanada K., Tanaka K.;
RT The novel adaptor protein, Mt1lp, and Vrp1p, a homolog of
RT Wiskott-Aldrich syndrome protein-interacting protein (WIP), may
RT antagonistically regulate type I myosins in Saccharomycetes
RT cerevisiae.
RL Genetics 160:923-934(2002).
RN [3]
RP SEQUENCE OF 775-1157 FROM N.A.
RA Pohl T.M., Aljinovic G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Involved in the regulation of actin cytoskeleton.
CC -1- SUBUNIT: Binds to the SH3 domains of the type I myosins MYO3 and

```

```

CC MYO5.
CC -1- SUBCELLULAR LOCATION: Co-localizes with cortical actin patches.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC in positions 729 and 732.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49295; CAA89311.1; ALT_FRAME.
DR EMBL; Z49296; CAA89312.1; ALT_INIT.
DR EMBL; AF373805; AAL57239.1; -.
DR SGD; S0003557; BBL1
DR GO; GO:0005857; C:actin cortical patch (sensu Saccharomycetes); IDA.
DR GO; GO:0017024; F:myosin I binding activity; IPI.
DR GO; GO:0030036; P:actin cytoskeleton organization and biogenesis; IGI.
DR GO; GO:0007011; P:regulation of cytoskeleton; IGI.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain; Coiled coil.
FT DOMAIN 5 69 SH3.
FT DOMAIN 234 301 COILED COIL (POTENTIAL).
FT DOMAIN 278 412 GLU-RICH.
FT DOMAIN 356 430 COILED COIL (POTENTIAL).
FT DOMAIN 674 830 PRO-RICH.
SQ SEQUENCE 1157 AA; 128296 MW; 88A58998B9CCE8ED CRC64;

Query Match
Best Local Similarity 6.1%; Score 123; DB 1; Length 1157;
Matches 94; Conservative 57; Mismatches 154; Indels 186; Gaps 21;

QY 11 PSAAEGRPTPTSPSLPSAQLPSHNPVPCQPLLDSTNSHQIMTDNDEEFPNS 70
DQ 707 PSVPSPAPPVPP--APPALSAPSPVPPVPPVSSAPPALSAPSIPIPPPTPAPPAPAP 764
QY 71 YLLRACSGPQQASSGPP-----NHHQSOSTLRPLP-----PP-----HNHT 107
DQ 765 LALPKHNEVEHVKSSAPLPVSEVHPMPNTAPLPAPPVPPATFEEDSEPTATHSHT 824
QY 108 L-----SHHH-----SSANSLNRN-----SLTNRRSIIHAPAPAP 137
DQ 825 APSPPPHQNVTAFTSMMSSTOQVRPTSVLSGAEKESRTLPPHPVPSLTNR-----PVDSF 878
QY 138 NDLATTPESVQL-----QDSWVLSNVP-----160
DQ 879 HESDTPKVASIRRTTHDVGESISNNVKIEFNAQERWINKSAPSAISNLKLNFLMEIDD 938
QY 161 -----LETRHFLFTSS---GSTPLFSSSSPGYPLTS-----GTVYTPPRL 200
DQ 939 HFISRLHQKVVVRDFYELFENYSQLRFSLTENSTSPKTVTLQERFSPVETQSARIL 998
QY 201 PRNTERKAFKLLKPKSKYCS-----WKCAALSAIAAALLAILLAYF 242
DQ 999 DEYA---QRFNAKVVEKSHSLNHSIGAKNFVSVQIVSEFKDEVIQPIGARTFGATLSY- 1054
QY 243 IVPWSLKNSSIDSGAEVQ-----RRVTOEVPVPGVFWRSQIHLSQPOFLKFNLSLGLKDALF 298
DQ 1055 -----KPEEGIEQLMSLQKIKPG-----DILVIRAKFAHKHKGKNEIL 1095
QY 299 GVIIRGLPPSH--AQYDFMERLDGKEKWSVSPRRRSIQTLYQNEAVFYQYLDVGLW 356
DQ 1096 NVGMDSAAPYSVWTDYDFT-----KNKFRVIENHEGK-----IIONS-----Y 1134
QY 357 HIAFYNDGDKD 367
DQ 1134

```

Db 1135 KLSHMSGKLG 1145

RESULT 10

GSRL_HUMAN

ID (GSRL_HUMAN STANDARD; PRT; 1509 AA.

AC QGNZM4;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Glioma tumor suppressor candidate region gene 1 protein.

GN GLTSCR1

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.

RX MEDLINE-20175430; PubMed-10708317;

RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,

RA Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,

RA Scheithauer B.W., Louis D.N., Jenkins R.B.;

RT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor

RT region";

RL Genomics 64:44-50(2000).

CC -1- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,

CC placenta, skeletal muscle, and pancreas, and at lower levels in

CC lung, liver, and kidney.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; AF182077; AAF62874.1; -

DR Genew; HGNC:4332; GLTSCR1.

DR MIM; 605690; -

DR DOMAIN 37 45 POLY-GLY.

FT DOMAIN 884 889 POLY-PRO.

FT DOMAIN 1214 1225 POLY-SER.

FT DOMAIN 1282 1286 POLY-PRO.

FT DOMAIN 1294 1304 POLY-PRO.

FT SEQUENCE 1509 AA; 152991 MW; 7C5144F443C6B821 CRC64;

SQ

Query Match 6.0%; Score 121; DB 1; Length 1509;

Best Local Similarity 23.3%; Pred. No. 1.2;

Matches 62; Conservative 26; Mismatches 94; Indels 84; Gaps 12;

QY 11 PSAEAGRIPPTSSPSLLPSAQL-PSHNPPVPVSCOMPLDNTSHQIMDTPDE---EF 66

Db 801 PTA-PGPQPPPLRQSQPEGPLPPAPHLPPSTSSAVASSSTSSRLPAPTSPDQLQF 859

QY 67 SPNSYLLRACSGPQQAASSGPPNHHQSQTLRPLPPPHN-----TLSSH 111

Db 860 PPS-----QGPKSPPTPTLHLVPEAAPPPPPPTFMVTTFPALPQPKALLERF 912

QY 112 H-----SSANSLNRLNRSQI-----HAPAPAPND 139

Db 913 HQVPSGIILONKAGAPAPQSTSLGPLTSPAASVLVSGQAPSGTPTAPSHAPAPAP-- 970

QY 140 LATTPEVSQLODSWVLSNVPLETRHFLFKTSSTPLFSSSPGYPPLTSGTVY----- 193

Db 971 MAATG-----LPPDLPAENKAFASNLPLNVAKAASSGPGRP--SGLOYESKLSG 1018

QY 194 -TPPPPLPNTTSRKAFKLKPKSKY 218

Db 1019 LKKPPTLPQ-----SKACFLEHLHKH 1040

RESULT 11

MUC2_HUMAN

ID MUC2_HUMAN STANDARD; PRT; 5179 AA.

AC Q02817; Q14878;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Mucin 2 precursor (Intestinal mucin 2).

GN MUC2 OR SMUC.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-94132002; PubMed-8300571;

RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;

RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.

RT Identification of the amino terminus and overall sequence similarity

RT to prepro-von Willebrand factor.";

RL J. Biol. Chem. 269:2440-2446(1994).

RN [2]

RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.

RC TISSUE-Colon;

RX MEDLINE-93016075; PubMed-1400449;

RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,

RA Kim Y.S.;

RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located

RT both upstream and downstream of its central repetitive region.";

RL J. Biol. Chem. 267:21375-21383(1992).

RN [3]

RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.

RX MEDLINE-91358717; PubMed-1885763;

RA Toribara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W.,

RA Petersen G.M., Kim Y.S.;

RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays

RT and polymorphism.";

RL J. Clin. Invest. 88:1005-1013(1991).

CC -1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND

CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A

CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS

CC AGENTS AT MUCOSAL SURFACES.

CC -1- SUBUNIT: MULTIMERIC.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,

CC BRONCHUS, CERVIX AND GALL BLADDER.

CC -1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR

CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).

CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND

CC VARIES AMONG DIFFERENT ALLELES.

CC -1- SIMILARITY: Contains 1 C-terminal cystine knot-like (CTCK) domain.

CC -1- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.

CC -1- SIMILARITY: Contains 2 VWFC domains.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; L21998; AAB95295.1; -

DR EMBL; M74027; AAA59875.1; -

DR EMBL; M94131; AAA59163.1; -

DR EMBL; M94132; AAA59164.1; -

DR PIR; A49963; A43932.

DR Genew; HGNC:7512; MUC2.

DR MIM; 158370; -

DR GO; GO:0005803; C:secretory vesicle; TAS.

DR InterPro; IPR006208; Cys_knot.

DR InterPro; IPR006207; Cys_knot_C.

DR InterPro; IPR006209; EGF_like.


```
Matches 91; Conservative 38; Mismatches 124; Indels 105; Gaps 20;
QY 4 AAMGRKSAE-AGRPDPTSSPSL-LPSAQLPSSHNPP-----VSCQM----- 46
Db 244 SATKQGGPTPSNGLPPGLSPSRVPSLQPPPLPPPPFPSPQSLMWSCSNLTS 303
QY 47 -----PLDNTSHQIMDTPDEEPSNLYLRACS-----GPQ-----QASSS 85
Db 304 GSTLRGSPNALLSMTSSNAALGPAMPYAPKLPSPALTPQPFQPSILANLMSSTI 363
QY 86 GPNHHSQSTL---RP-PLPPPHNHTLSHHSSANSNLSNRSHQIHAPAPANDLA 141
Db 364 KTPQGLHMGALPASNGPSPYRPEKL-----SSPGLPOOSFTPQCSLRSLRPTNLS 418
QY 142 TPYESVLQDSWVLSNVPLETRHFLK-TSSGSTPLFS-----SSSPG---YPLTSG 190
Db 419 QOQOQOQOQO---QANV-----IFKPISSNSKTLNMQOQWASSPGATEFTFG 467
QY 191 T-----VYTPPRLLP-RNTFSRK---AFKLKPKSKYCSWKCAALSAIAALLAIL 239
Db 468 NTKPLSHFVSEPCPKMPSMPTTSRQPSLLHYLQOPT-----PTQASSATASATATL- 521
QY 240 AFYIVPNSLKNISDSGEAEVGRVTOEYPPGVFWRSQIHISQPFKFNISLKGKAL 297
Db 522 -----QLQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 560

RESULT 13
BCAL_RAT
ID BCAL_RAT STANDARD; PRT: 968 AA.
AC Q63767; Q63766;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CRK-associated substrate (P130CAS) (Breast cancer anti-estrogen
DE resistance 1 protein).
GN BCAR1 OR CRKAS OR CAS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Fibroblast;
RX MEDLINE=94349922; PubMed=8070403;
RA Sakai R., Iwamatsu A., Hirano N., Ogawa S., Tanaka T., Mano H.,
RA Yazaki Y., Hirai H.;
RT "A novel signaling molecule, p130, forms stable complexes in vivo with
RT v-Crk and v-Src in a tyrosine phosphorylation-dependent manner.";
RL EMBO J. 13:3748-3756(1994).
RN [2]
RP TYROSINE PHOSPHORYLATION BY FOCAL ADHESION KINASE.
RX MEDLINE=98030588; PubMed=9360983;
RA Tachibana K., Urano T., Fujita H., Ohashi Y., Kamiguchi K., Iwata S.,
RA Hirai H., Morimoto C.;
RT "Tyrosine phosphorylation of Crk-associated substrates by focal
RT adhesion kinase. A putative mechanism for the integrin-mediated
RT tyrosine phosphorylation of Crk-associated substrates.";
RL J. Biol. Chem. 272:29083-29090(1997).
CC -1- FUNCTION: DOCKING PROTEIN WHICH PLAYS A CENTRAL COORDINATING ROLE
CC FOR TYROSINE-KINASE-BASED SIGNALING RELATED TO CELL ADHESION.
CC IMPLICATED IN INDUCTION OF CELL MIGRATION (BY SIMILARITY).
CC -1- SUBUNIT: FORMS COMPLEXES IN VIVO WITH FOCAL ADHESION KINASE 1,
CC ADAPTER PROTEIN CRKL AND LYN KINASE. CAN HETERODIMERIZE WITH CASL.
CC -1- SUBCELLULAR LOCATION: FOCAL ADHESIONS AND STRESS FIBERS.
CC UNPHOSPHORYLATED FORM LOCALIZES IN THE CYTOPLASM AND CAN MOVE TO
CC THE MEMBRANE UPON TYROSINE PHOSPHORYLATION.
CC -1- Alternative products:
CC Name=Long;
CC IsoId=Q63767-1; Sequence=Displayed;
CC Name=Short;
```

```
CC IsoId=Q63767-2; Sequence=VSP_004135;
CC Note-No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHER EXPRESSION IN LUNG,
CC INTESTINE AND TESTIS.
CC -1- DOMAIN: CONTAINS A CENTRAL DOMAIN (SUBSTRATE DOMAIN) CONTAINING
CC MULTIPLE POTENTIAL SH2-BINDING SITES AND A C-TERMINAL DOMAIN
CC CONTAINING A DIVERGENT HELIX-LOOP-HELIX (HLH) MOTIF. THE SH2-
CC BINDING SITES PUTATIVELY BIND CRK, NCK AND ABL SH2 DOMAINS. THE
CC HLH MOTIF IS ABSOLUTELY REQUIRED FOR THE INDUCTION OF PSEUDOHYPHAL
CC GROWTH IN YEAST AND MEDIATES HETERODIMERIZATION WITH CASL.
CC -1- DOMAIN: A SERINE-RICH REGION PROMOTES ACTIVATION OF THE SERUM
CC RESPONSE ELEMENT (SRE).
CC -1- DOMAIN: THE SH3 DOMAIN IS NECESSARY FOR THE LOCALIZATION OF THE
CC PROTEIN TO FOCAL ADHESIONS AND INTERACTS WITH ONE PROLINE-RICH
CC REGION OF FOCAL ADHESION KINASE 1.
CC -1- PTM: FOCAL ADHESION KINASE 1 PHOSPHORYLATES THE PROTEIN AT THE
CC YDYVHL MOTIF. THE SRC FAMILY KINASES ARE RECRUITED TO THE
CC PHOSPHORYLATED SITES AND CAN PHOSPHORYLATE OTHER TYROSINE
CC RESIDUES. TYROSINE PHOSPHORYLATION IS TRIGGERED BY INTEGRIN
CC MEDIATED ADHESION OF CELLS TO THE EXTRACELLULAR MATRIX.
CC -1- DISEASE: APPEARS TO HAVE A CENTRAL FUNCTION IN TRANSFORMATION OF
CC SOME CELL TYPES.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -1- SIMILARITY: BELONGS TO THE CAS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D29766; BAA06169.1; -
CC EMBL: D29766; BAA06170.1; -
CC HSSP: P07751; IBK2.
CC InterPro: IPR001452; SH3.
CC Pfam: PF00018; SH3; 1.
CC PRINTS: PR00452; SH3DOMAIN.
CC ProDom: PD000066; SH3; 1.
CC SMART: SM00326; SH3; 1.
CC PROSITE: PS50002; SH3; 1.
CC DR Phosphorylation; SH3 domain; SH3-binding; Cell adhesion;
CC KW Alternative splicing.
CC FT DOMAIN 97 159 SH3.
CC FT DOMAIN 168 181 PRO-RICH.
CC FT DOMAIN 213 514 SUBSTRATE FOR KINASES.
CC FT DOMAIN 520 712 SER-RICH.
CC FT SITE 733 741 SH3-BINDING (POTENTIAL).
CC FT DOMAIN 844 894 DIVERGENT HELIX-LOOP-HELIX MOTIF.
CC FT VARSPLIC 5 98 Missing (in isoform Short).
CC FT /FTID=VSP_004135
CC SQ SEQUENCE 968 AA; 104262 MW; E861641BF6D68D377 CRC64;
Query Match 6.0%; Score 120.5; DB 1; Length 968;
Best Local Similarity 22.9%; Pred. No. 0.76;
Matches 80; Conservative 35; Mismatches 84; Indels 151; Gaps 24;
QY 9 RKPSAAGRPPTTSPSLLPSAQLPSS-HNP-PPVSCQMPLLDSNTSHQIMDTPDEEF 66
Db 163 KRPAAPG--PGPATPPQ--PQPSLPGQVHTPPASQVSPMLP--TAYO---PQPD--- 210
QY 67 SPNSYLLRACSGPQQAASSGP-PNHHSQSFTLRPLPPPHNHTLSH-----HHSANSNLR 120
Db 211 --NVILVPPSKTQOGLYQAPGNPQFS-----PPAKQTSTFSKQTPHHS----- 254
QY 121 NSLTNRRSQIHAPAPNDLATTPESVQLQDSWVLSNVPLETRHFLFTKTSSTPLFSS 180
Db 255 -----FPSPA-TDLYQVP----- 266
QY 181 SSPGYPL-----TSGT---VYTPPRLPNTFSRKAFKLKPKSKYCSWKCAALSAI 229
Db 267 PGPSPAQDIYQVPPSAGTGHDIYQVPPSL-----DTRSWEGTKPP----- 307
```


DR EMBL; AF445387; AAN75849.1; JOINED.
DR EMBL; AF445388; AAN75849.1; JOINED.
DR EMBL; AF445390; AAN75849.1; JOINED.
DR EMBL; AF445391; AAN75849.1; JOINED.
DR EMBL; AF445392; AAN75849.1; JOINED.
DR EMBL; AF445393; AAN75849.1; JOINED.
DR EMBL; AF445394; AAN75849.1; JOINED.
DR EMBL; AF445395; AAN75849.1; JOINED.
DR EMBL; AF445396; AAN75849.1; JOINED.
DR EMBL; AF445397; AAN75849.1; JOINED.
DR EMBL; AF445398; AAN75849.1; JOINED.
DR EMBL; AF445399; AAN75849.1; JOINED.
DR EMBL; AF445400; AAN75849.1; JOINED.
DR EMBL; AF445401; AAN75849.1; JOINED.
DR EMBL; AF445402; AAN75849.1; JOINED.
DR EMBL; AF445403; AAN75849.1; JOINED.
DR EMBL; AF445404; AAN75849.1; JOINED.
DR EMBL; AF445405; AAN75849.1; JOINED.
DR EMBL; AF445406; AAN75849.1; JOINED.
DR EMBL; AF445407; AAN75849.1; JOINED.
DR EMBL; AF445408; AAN75849.1; JOINED.
DR EMBL; AF445409; AAN75849.1; JOINED.
DR EMBL; AF445410; AAN75849.1; JOINED.
DR EMBL; AF445411; AAN75849.1; JOINED.
DR EMBL; AF445412; AAN75849.1; JOINED.
DR EMBL; AF445413; AAN75850.1; JOINED.
DR EMBL; AF445385; AAN75850.1; JOINED.
DR EMBL; AF445386; AAN75850.1; JOINED.
DR EMBL; AF445387; AAN75850.1; JOINED.
DR EMBL; AF445388; AAN75850.1; JOINED.
DR EMBL; AF445390; AAN75850.1; JOINED.
DR EMBL; AF445391; AAN75850.1; JOINED.
DR EMBL; AF445392; AAN75850.1; JOINED.
DR EMBL; AF445393; AAN75850.1; JOINED.
DR EMBL; AF445394; AAN75850.1; JOINED.
DR EMBL; AF445395; AAN75850.1; JOINED.
DR EMBL; AF445397; AAN75850.1; JOINED.
DR EMBL; AF445398; AAN75850.1; JOINED.
DR EMBL; AF445399; AAN75850.1; JOINED.
DR EMBL; AF445399; AAN75850.1; JOINED.
DR EMBL; AF445400; AAN75850.1; JOINED.
DR EMBL; AF445401; AAN75850.1; JOINED.
DR EMBL; AF445402; AAN75850.1; JOINED.
DR EMBL; AF445403; AAN75850.1; JOINED.
DR EMBL; AF445404; AAN75850.1; JOINED.
DR EMBL; AF445405; AAN75850.1; JOINED.
DR EMBL; AF445406; AAN75850.1; JOINED.
DR EMBL; AF445407; AAN75850.1; JOINED.
DR EMBL; AF445408; AAN75850.1; JOINED.
DR EMBL; AF445409; AAN75850.1; JOINED.
DR EMBL; AF445410; AAN75850.1; JOINED.
DR EMBL; AF445411; AAN75850.1; JOINED.
DR EMBL; AF445412; AAN75850.1; JOINED.
DR EMBL; AC007684; AAF19240.1; JOINED.
DR HSSP; P24941; 1838.
DR Genew; HGNC:6865; MAP4K3.
DR MIM; 604921; -.
DR InterPro; IPR001180; Citron.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR01245; Tyr_pkinase.
DR Pfam; PF00780; CNH; 1.
DR Pfam; PF00069; pkinase; 1.
DR -ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00036; CNH; 1.
DR SMART; SM00220; S-TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
KW ATP-binding; Transferase; Serine/threonine-protein kinase;
KW Alternative splicing.
FT DOMAIN 16 273 PROTEIN_KINASE.
FT DOMAIN 562 874 CNH.
FT NP_BIND 22 30 ATP (BY SIMILARITY).

FT BINDING 48 48 ATP.
FT ACT_SITE 136 136 BY SIMILARITY.
FT VARSPLIC 1 12 MNPFDLSRRNP -> MA (in isoform 2).
FT VARSPLIC 352 372 /FTId=VSP_007052.
FT VARSPLIC 352 372 Missing (in isoform 3).
FT MUTAGEN 48 48 /FTId=VSP_007053.
FT CONFLICT 392 392 K->E: LOSS OF KINASE ACTIVITY AND ABILITY
FT CONFLICT 392 392 TO ACTIVATE JNK FAMILY.
FT SEQUENCE 894 AA; 101315 MW; 5EB77BBB34E5B733 CRC64;
Query Match 5.88; Score 117.5; DB 1; Length 894;
Best Local Similarity 23.18; Pred. No. 1.1;
Matches 69; Conservative 30; Mismatches 93; Indels 107; Gaps 15;
QY 31 AQLPSSHNPPVSCQMPLLDSNTSHQIMDTNPDEEFPNSYLLRACSGPOQASSSGPPNH 90
Db 429 AKIPPLPPPKPSIFIP-----QEMHSTEDE---NOGTIKRC-----PMSGSPAK 470
QY 91 HSQSTLRPLP---PPHNHTLSHHSSANSLN-----RNSLTNRSSQIHAPA 134
Db 471 PSQVPPPPPPRLPPHKKPVALGNGMSSFQNGERDGLCOQOQHRTNLSRKEKDVPK 530
QY 135 PAPNDLATPE-----SVQLQDSWV-----LNSNV 159
Db 531 PISNGLPPTPKVHMGACFSKVFNGCPLKHCASSWINPDRDQYLIFGAEEGYITLNE 590
QY 160 PLET-----RHLEKTSSTGSLPFLSSSPGY-----PLTSGTVYTP---195
Db 591 LHETSMELPPRRTWLYVNNCLLSISGRASQLYSHNPLGLFDYARQMKLPVPAHK 650
QY 196 -PPRLLPRTFSRKAFKLRKPK-SKYCSWKCAAL-----SAIAAAALLAILLAYFTVP 245
Db 651 LPDRILPRK-FSVSA---KIPETKWCQ-KCCVVRNPVTGHKYLCCGALQTSIVLLEWYEP 704

Search completed: August 26, 2003, 14:31:13
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:29:03 ; Search time 44 Seconds
(without alignments)
2234.501 Million cell updates/sec

Title: US-09-584-411c-22

Perfect score: 2017

Sequence: 1 MLHANKGRPSAAGRP...NDGKKEWSEFNTVLDGTI 381

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|--------|---------------------|
| 1 | 1889.5 | 93.7 | 2764 | 11 | Q9WTS5 | Q9WTS5 mus musculus |
| 2 | 1889.5 | 93.7 | 2765 | 11 | Q9R1K2 | Q9R1K2 rattus norv |
| 3 | 1832.5 | 90.9 | 831 | 13 | Q9PU49 | Q9PU49 gallus gall |
| 4 | 1820.5 | 90.3 | 2802 | 13 | Q9DER5 | Q9DER5 gallus gall |
| 5 | 975.5 | 48.4 | 2715 | 11 | Q9WTS6 | Q9WTS6 mus musculus |
| 6 | 917 | 45.5 | 2590 | 13 | Q9W7R4 | Q9W7R4 brachydanio |
| 7 | 853 | 42.3 | 2825 | 11 | O70465 | O70465 mus musculus |
| 8 | 830 | 41.2 | 2824 | 13 | Q9W7R3 | Q9W7R3 brachydanio |
| 9 | 793.5 | 39.3 | 2771 | 11 | Q9WTS7 | Q9WTS7 mus musculus |
| 10 | 790.5 | 39.2 | 560 | 4 | P98202 | P98202 homo sapien |
| 11 | 770 | 38.2 | 442 | 4 | Q9NZJ2 | Q9NZJ2 homo sapien |
| 12 | 745.5 | 37.0 | 2705 | 13 | Q9W6V6 | Q9W6V6 gallus gall |
| 13 | 739 | 36.6 | 536 | 11 | O8C8D2 | O8C8D2 mus musculus |
| 14 | 739 | 36.6 | 2725 | 4 | O9UK24 | O9UK24 homo sapien |
| 15 | 737.5 | 36.6 | 2731 | 11 | Q9WTS4 | Q9WTS4 mus musculus |
| 16 | 461.5 | 22.9 | 2346 | 11 | Q9JLC1 | Q9JLC1 mus musculus |

| | | | | | | |
|----|-------|------|------|----|--------|--------------------|
| 17 | 427 | 21.2 | 159 | 4 | 075999 | 075999 homo sapien |
| 18 | 207.5 | 10.3 | 2731 | 5 | Q9VNU6 | Q9VNU6 drosophila |
| 19 | 207.5 | 10.3 | 2731 | 5 | O61307 | O61307 drosophila |
| 20 | 205.5 | 10.2 | 2731 | 5 | O18366 | O18366 drosophila |
| 21 | 186.5 | 9.2 | 2515 | 5 | O24551 | O24551 drosophila |
| 22 | 177 | 8.8 | 1111 | 10 | Q9FH74 | Q9FH74 arabidopsis |
| 23 | 177 | 8.8 | 1113 | 10 | O8LPR4 | O8LPR4 arabidopsis |
| 24 | 156.5 | 7.8 | 528 | 6 | Q29071 | Q29071 sus scrofa |
| 25 | 154.5 | 7.7 | 3004 | 5 | Q24550 | Q24550 drosophila |
| 26 | 154.5 | 7.7 | 3004 | 5 | Q9VYN8 | Q9VYN8 drosophila |
| 27 | 144 | 7.1 | 309 | 4 | Q9H7C7 | Q9H7C7 homo sapien |
| 28 | 141 | 7.0 | 2531 | 5 | Q8MP22 | Q8MP22 caenorhabdi |
| 29 | 141 | 7.0 | 2560 | 5 | Q21980 | Q21980 caenorhabdi |
| 30 | 137.5 | 6.8 | 325 | 4 | Q8NAK1 | Q8NAK1 homo sapien |
| 31 | 136.5 | 6.8 | 1895 | 5 | Q9VWS5 | Q9VWS5 drosophila |
| 32 | 135 | 6.7 | 1466 | 10 | Q94HW7 | Q94HW7 arabidopsis |
| 33 | 134.5 | 6.7 | 577 | 4 | O8N811 | O8N811 homo sapien |
| 34 | 132.5 | 6.6 | 2703 | 5 | Q9VEG7 | Q9VEG7 drosophila |
| 35 | 132.5 | 6.6 | 2715 | 5 | O61603 | O61603 drosophila |
| 36 | 132.5 | 6.6 | 2716 | 5 | Q8IN94 | Q8IN94 drosophila |
| 37 | 132 | 6.5 | 421 | 4 | O15329 | O15329 homo sapien |
| 38 | 132 | 6.5 | 1268 | 11 | O63623 | O63623 rattus norv |
| 39 | 132 | 6.5 | 1466 | 10 | Q94HW2 | Q94HW2 arabidopsis |
| 40 | 131 | 6.5 | 1892 | 5 | O57007 | O57007 leishmania |
| 41 | 129.5 | 6.4 | 731 | 10 | O65530 | O65530 arabidopsis |
| 42 | 129 | 6.4 | 383 | 5 | Q9Y075 | Q9Y075 leishmania |
| 43 | 129 | 6.4 | 1330 | 10 | Q9SXQ2 | Q9SXQ2 arabidopsis |
| 44 | 129 | 6.4 | 1456 | 10 | O92T94 | O92T94 arabidopsis |
| 45 | 128.5 | 6.4 | 518 | 3 | Q8NJ53 | Q8NJ53 candida alb |

ALIGNMENTS

RESULT 1

| | | |
|--------|--|---------------|
| Q9WTS5 | PRELIMINARY; | PRT; 2764 AA. |
| AC | Q9WTS5; | |
| DT | 01-NOV-1999 (TREMBLrel. 12, Created) | |
| DT | 01-NOV-1999 (TREMBLrel. 12, Last sequence update) | |
| DT | 01-MAR-2003 (TREMBLrel. 23, Last annotation update) | |
| DE | Ten-m2. | |
| GN | OD22 OR TEN-M2. | |
| OS | Mus musculus (Mouse). | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. | |
| OX | NCBI_taxID=10090; | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RC | STRAIN=Balb/c; TISSUE=Brain; | |
| RA | Ohashi T., Zhou X., Feng K., Richter B., Moergelin M., Perez M.T., | |
| RA | Su W., Chiquet-Ehrismann R., Rauch U., Faessler R.; | |
| RT | "Mouse Ten-m/Odz is a new family of dimeric type II transmembrane | |
| RT | proteins expressed in many tissues." | |
| RL | J. Cell Biol. 0:0-0(1999). | |
| DR | EMBL; AB025411; BAA77397.1; - | |
| DR | HSP; P35555; 1ENN. | |
| DR | MGD; MGI:1345184; Odz2. | |
| DR | InterPro; IPR006209; EGF_like. | |
| DR | InterPro; IPR006210; IEGF. | |
| DR | InterPro; IPR002049; Laminin_EGF. | |
| DR | InterPro; IPR001258; NHL. | |
| DR | InterPro; IPR006530; YD. | |
| DR | Pfam; PF00008; EGF; 4. | |
| DR | Pfam; PF01436; NHL; 6. | |
| DR | PRINTS; PRO0011; EGF_LAMININ. | |
| DR | SMART; SM00181; EGF; 5. | |
| DR | TIGRFAMs; TIGR01643; YD_repeat_2x; 7. | |
| DR | PROSITE; PS00022; EGF_1; 8. | |
| DR | PROSITE; PS01186; EGF_2; 7. | |
| KW | EGF-like domain. | |
| SQ | SEQUENCE 2764 AA; 306464 MW; 73BA3D916D0F0344 CRC64; | |

Query Match 93.7%; Score 1889.5; DB 11; Length 2764;
 Best Local Similarity 88.5%; Pred. No. 6.3e-161;
 Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;

QY 10 KPSAAGRIPIPTSSPSLLPSAQLPSSHNPPVPSQMPLLDSNTSHQIMDTPDEEFSN 69
 DB 162 KDDNGRIPIPTSSPSLLPSAQLPSSHNPPVPSQMPLLDSNTSHQIMDTPDEEFSN 221

QY 70 SYLLRACSGPQQAASSGPPNHHSSQSLRPPLPPPHNHTLSHHSSANSLSNRSLTNRRSQ 129
 DB 222 SYLLRACSGPQQAASSGPPNHHSSQSLRPPLPPPHNHTLSHHSSANSLSNRSLTNRRSQ 281

QY 130 IHAPAPAPNDLATTPEVOLQDSWLVNSVPLETRHFLFKTSSTGSPFLSSSSPGYPLTS 189
 DB 282 IHAPAPAPNDLATTPEVOLQDSWLVNSVPLETRHFLFKTSSTGSPFLSSSSPGYPLTS 341

QY 190 GVIYTPPPRLPNTSRAFKLKPKSKYCSWKCAALSAAIAALLAILLAYFI----- 243
 DB 342 GVIYTPPPRLPNTSRAFKLKPKSKYCSWKCAALSAAIAALLAILLAYFIAMHLLG 401

QY 244 -----VPWSLKNSSIDSGAEVGRRTQEV 268
 DB 402 LNWQLPADGHTFNNGVTRGLPGNDVAVTPSGGKVPWSLKNSSIDSGAEVGRRTQEV 461

QY 269 PGVFWRSQIHSQPOFLKFNISLGDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV 328
 DB 462 PGVFWRSQIHSQPOFLKFNISLGDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV 521

QY 329 ESPRRRSQITLVQNEAVFQVLDVGLHFLAFYNDGKEMVSNFTVWLD 378
 DB 522 ESPRRRSQITLVQNEAVFQVLDVGLHFLAFYNDGKEMVSNFTVWLD 571

RESULT 2
 Q9RIK2 ID Q9RIK2 PRELIMINARY; PRT; 2765 AA.
 AC Q9RIK2
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Neurestin alpha.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP "Neurestin: putative transmembrane molecule implicated in neuronal development."
 RL Dev. Biol. 212:165-181(1999).
 DR EMBL; AF086607; AAD47383.1; .
 DR HSSP; P00750; ITPG.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001258; NHL.
 DR InterPro; IPR006530; YD.
 DR Pfam; PF00008; EGF_4.
 DR Pfam; PF01436; NHL; 6.
 DR SMART; SM00181; EGF; 5.
 DR TIGRfams; TIGR01643; YD_repeat_2x; 7.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 7.
 KW EGF-like domain.
 SQ SEQUENCE 2765 AA; 306568 MW; 6748D70D5FFD8F0E CRC64;

Query Match 93.7%; Score 1889.5; DB 11; Length 2765;
 Best Local Similarity 88.5%; Pred. No. 6.3e-161;
 Matches 363; Conservative 1; Mismatches 5; Indels 41; Gaps 1;

QY 10 KPSAAGRIPIPTSSPSLLPSAQLPSSHNPPVPSQMPLLDSNTSHQIMDTPDEEFSN 69

DB 162 KDDNGRIPIPTSSPSLLPSAQLPSSHNPPVPSQMPLLDSNTSHQIMDTPDEEFSN 221

QY 70 SYLLRACSGPQQAASSGPPNHHSSQSLRPPLPPPHNHTLSHHSSANSLSNRSLTNRRSQ 129
 DB 222 SYLLRACSGPQQAASSGPPNHHSSQSLRPPLPPPHNHTLSHHSSANSLSNRSLTNRRSQ 281

QY 130 IHAPAPAPNDLATTPEVOLQDSWLVNSVPLETRHFLFKTSSTGSPFLSSSSPGYPLTS 189
 DB 282 IHAPAPAPNDLATTPEVOLQDSWLVNSVPLETRHFLFKTSSTGSPFLSSSSPGYPLTS 341

QY 190 GVIYTPPPRLPNTSRAFKLKPKSKYCSWKCAALSAAIAALLAILLAYFI----- 243
 DB 342 GVIYTPPPRLPNTSRAFKLKPKSKYCSWKCAALSAAIAALLAILLAYFIAMHLLG 401

QY 244 -----VPWSLKNSSIDSGAEVGRRTQEV 268
 DB 402 LNWQLPADGHTFNNGVTRGLPGNDVAVTPSGGKVPWSLKNSSIDSGAEVGRRTQEV 461

QY 269 PGVFWRSQIHSQPOFLKFNISLGDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV 328
 DB 462 PGVFWRSQIHSQPOFLKFNISLGDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV 521

QY 329 ESPRRRSQITLVQNEAVFQVLDVGLHFLAFYNDGKEMVSNFTVWLD 378
 DB 522 ESPRRRSQITLVQNEAVFQVLDVGLHFLAFYNDGKEMVSNFTVWLD 571

RESULT 3
 Q9PU49 ID Q9PU49 PRELIMINARY; PRT; 831 AA.
 AC Q9PU49
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Teneurin-2 protein.
 GN TEN2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20057765; PubMed-10588872;
 RA Rubin B.P., Tucker R.P., Martin D., Chiquet-Ehrismann R.;
 RT "Teneurins: a novel family of neuronal cell surface proteins in
 RT vertebrates, homologous to the Drosophila pair-rule gene product Ten-
 RT m.";
 RL Dev. Biol. 216:195-209(1999).
 DR EMBL; AJ245711; CAB57257.1; .
 DR HSSP; P00750; ITPG.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 4.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 5.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01186; EGF_2; 6.
 KW EGF-like domain.
 SQ SEQUENCE 831 AA; 90359 MW; F166547445EAF7D2 CRC64;

Query Match 90.9%; Score 1832.5; DB 13; Length 831;
 Best Local Similarity 85.1%; Pred. No. 1.8e-156;
 Matches 349; Conservative 9; Mismatches 11; Indels 41; Gaps 1;

QY 10 KPSAAGRIPIPTSSPSLLPSAQLPSSHNPPVPSQMPLLDSNTSHQIMDTPDEEFSN 69
 DB 162 KDDNGRIPIPTSSPSLLPSAQLPSSHNPPVPSQMPLLDSNTSHQIMDTPDEEFSN 221

QY 70 SYLLRACSGPQQAASSGPPNHHSSQSLRPPLPPPHNHTLSHHSSANSLSNRSLTNRRSQ 129

Db 256 TGTGTTPLFSTATPGYTASGSVYSPTRPLPNTLSRSFAFKKSKYCSWRCALCAV 315
 Qy 230 AAALLAILLAYFI-----VPWSL----- 248
 Db 316 GVSVLLAILLSYFIAMHFLGNLHQTENDTFENGKVSNDTPTNTVSLPSGDNGKLGG 375
 Qy 249 ---KNSIDSCEAEVGRRTQEPVPGVFRSQTHISQPLKFNISLGDALGVYIRRG 305
 Db 376 FTHTNTIDSGEDIGIRRAIQEYPPGIFWRSOLFDPQPLKFNISLQKDALIGVYGRKG 435
 Qy 306 LPSPHAYDPMERLDG-----KSKWSVESPRRRRTTOTLVQNEAVFQYLDVGLWHLAF 360
 Db 436 LPSPHAYDPMERLDG-----KSKWSVESPRRRRTTOTLVQNEAVFQYLDVGLWHLAF 360
 Qy 361 YNDGDKKEMVFNFTVLDGTI 381
 Db 496 YNDGKNPEQVFNFTVIESV 516

RESULT 6

Q9W7R4 ID Q9W7R4 PRELIMINARY; PRT; 2590 AA.
 AC Q9W7R4;
 DT 01-NOV-1999 (TEMBLrel. 12, Created)
 DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Ten-m3.
 GN OD23 OR TENM3 OR TEN-M3.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OC NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99425191; PubMed-10495292;
 RA Mieda M., Kikuchi Y., Hirate Y., Aoki M., Okamoto H.;
 RT "Compartmentalized expression of zebrafish ten-m3 and ten-m4,
 RT homologues of the Drosophila tenm /odd Oz gene, in the central nervous
 RT system.";
 RL Mech. Dev. 87:223-227(1999).
 DR EMBL; AB026979; BAA81892.1; -.
 DR HSSP; P10968; 7WGA.
 DR ZFIN; ZDB-GENE-990714-19; odt3.
 DR InterPro; IPR006210; EGF-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001258; NHL.
 DR InterPro; IPR006530; YD.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF01436; NHL; 6.
 DR SMART; SM00181; EGF; 6.
 DR TIGRFAMs; TIGR01643; YD_repeat_2x; 7.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 7.
 KW EGF-like domain.
 SQ SEQUENCE 2590 AA; 288586 MW; 597592866210148D CRC64;

Query Match 45.5%; Score 917; DB 13; Length 2590;
 Best Local Similarity 46.5%; Pred. No. 3.2e-73;
 Matches 190; Conservative 61; Mismatches 94; Indels 64; Gaps 9;
 Qy 23 SPSSLPSAQLPSSNPPPV-----SCQPLLDNNTSHQIMDNDDEEFPNSVLLRACSG 78
 Db 6 SPSPV-----TEHSHQPPSPNLDNQSSILSNATTQAVQSDSEETAVLYRPVYTPA 60
 Qy 79 PQQASSGPPNHH--SQSLRPLPPPHNHTLHHSSANSLSNRSLNRSQIHA-PAP 135
 Db 61 PSHSCNEQSNHQOQCSLT-PPVPPPHK-----QPSVTALMHNLSLRNNSVAPPAA 114
 Qy 136 APNDLATTPEVQLODSWLVNSVPLETRHFLKTSSTGTPFLFSSSSSGYPLTSGVYTP 195
 Db 115 LPALQTTPEVQLODSWLVNSVPLETRHFLKTSSTGTPFLFSSSSSGYPLTSGVYTP 174

Qy 196 PPRLPNTSRAFKLKKPKSKYCSWKAALSAIAAALLAILLAY-----FIVPWSL- 248
 Db 175 PTRPLPNTLSRAFAFKKSKYCSWRCALTASAMAVSILLSVLLCYCIAMHFLGNLQ 234
 Qy 249 -----KNSIDSCEAEVGRRTQEPVPGVFR 274
 Db 235 ETEGYAFENGQVKSSTATNAVLTALSTENKVFYQENNTIDTGEVDVGRRAVDVPPGTFW 294
 Qy 275 RSQIHISQPLKFNISLGDALGVYIRRGVFRSQTHISQPLKFNISLGDALGVYIRRG 329
 Db 295 RTQOLFDPQPLKFNISVQGRKGLPETHQYDFVLLDGSRLIAEKKRGLVE 354
 Qy 330 SPERRSIQPLVQNEAVFQYLDVGLWHLAFYNDGDKKEMVFNFTVLD 378
 Db 355 VEGAARKARSVNVHEAEFIRFLDSGTWHLAFYNDGKNAEQVSYNTIID 403

RESULT 7

O70465 ID O70465 PRELIMINARY; PRT; 2825 AA.
 AC O70465;
 DT 01-AUG-1998 (TEMBLrel. 07, Created)
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE DOC4.
 GN NRGI OR OD24 OR DOC4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=NIH/Swiss;
 RX MEDLINE-98315054; PubMed-9649432;
 RA Wang X.Z., Kuroda M., Sok J., Batchvarova N., Kimmel R., Chung P.,
 RA Zinszner H., Ron D.;
 RT "Identification of novel stress-induced genes downstream of chop.";
 RL EMBO J. 17:3619-3630(1998).
 DR EMBL; AF059485; AAC31807.1; -.
 DR HSSP; P16109; 1FSB.
 DR MGD; MGI:96083; Nrql.
 DR MGD; MGI:2447063; Otd4.
 DR InterPro; IPR006210; EGF-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001258; NHL.
 DR InterPro; IPR006530; YD.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF01436; NHL; 6.
 DR SMART; SM00181; EGF; 6.
 DR TIGRFAMs; TIGR01643; YD_repeat_2x; 7.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 7.
 KW EGF-like domain.
 SQ SEQUENCE 2825 AA; 313483 MW; A120D98080886032 CRC64;

Query Match 42.3%; Score 853; DB 11; Length 2825;
 Best Local Similarity 40.1%; Pred. No. 2.1e-67;
 Matches 193; Conservative 55; Mismatches 111; Indels 122; Gaps 12;
 Qy 12 SAAGAPRIPPTSPSLLPSAQLPSSNPPPV-----NHSQSLRPLPPPHNHTLHHH--SSANSLSNRSLN 70
 Db 161 NTEGTAPL-HCSSASSTPEQSPSPPPSPANESQRRLLGNGVAQTPDSDSEEEFVPS 219
 Qy 71 YLIRACSGPQQAASSGPP-----NHSQSLRPLPPPHNHTLHHH--SSANSLSNRSLN 125
 Db 220 FLVKSSASLGVAAVDHPSSLQNHPLRT--PPPLPHANTPHQHAASINSUNRGFT 277
 Qy 126 RRSQIHAPAPND-----LATTPEVQLODSWLVNSVPLETR----- 164
 Db 278 RSN-----PSPAPTDHSLSGEPAPAGSAQEPHQAQDNVLSKIPVETRNLGKQPLGTWQD 333
 Qy 165 -----HFLFKYSSGSTPLFSSSSSGYPLTSGVYTPPPRLPRLPNT 204

```

Db 334 NLIEMDIASRRDYGAYSDGHEFFK-PCGTSPLFCTTSPGYDLTSTVSPRPRLPRST 392
QY 205 FSRKAPLKKPKSKYCWKAALSAIAAALLAILLAYFI-----VPWSL----- 248
Db 393 FSRPAFLKPKSKYCNWKAALSAIALISATVILLAYFVAMHLFGLNHLQPMEGOMQMY 452
QY 249 -----EKWSVSPRRRISIOFLVQNEAVFYQYLDVGLWHLAFYNDGDKKEMVSFNTVVL 262
Db 453 EYTEDTASWPVPTDVSLYPSGGTGLETPDRKGGAEGKPSLFPEDSFIDSGEIDVGR 512
QY 263 RYTOEVPVGVWRWSOIHSIQPKFNISLGRDGLFVYIRRGLPSPSHQAQYDFMERLDGK 322
Db 513 RASQKIPGCTFWRSQVFDHPVHLKFNLSGRALVYIGRGLPSPHTQLDFVELLDG 572
QY 323 -----EKWSVSPRRRISIOFLVQNEAVFYQYLDVGLWHLAFYNDGDKKEMVSFNTVVL 377
Db 573 RLUTQEARSLQEPQSRGCPVPSPSHETGFIQYLDGSIWHLAFYNDGKESEVVSFLTITAI 632
QY 378 D 378
Db 633 E 633

RESULT 8
QW7R3
ID QW7R3 PRELIMINARY; PRT; 2824 AA.
AC QW7R3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NAR-2003 (TREMBlrel. 23, Last annotation update)
DE Ten-m4.
GN NRGL OR TENM4 OR TEN-M4.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99425191; PubMed=10495292;
RA Mieda M., Kikuchi Y., Hirate Y., Aoki M., Okamoto H.;
RT "Compartmentalized expression of zebrafish ten-m3 and ten-m4,
RT homologues of the Drosophila tenm/odd oz gene, in the central nervous
RT system.";
RL Mech. Dev. 87:223-227(1999).
DR EMBL; AB026980; BAA81893.1; -.
DR HSSP; P10959; 1WGT.
DR ZFIN; ZDB-GENE-990714-20; nrgl.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001258; NHL.
DR InterPro; IPR006530; YD.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF01436; NHL; 5.
DR PRINTS; PRO0011; EGF_LAMININ.
DR SMART; SM00181; EGF; 6.
DR TIGRfams; TIGR01643; YD_repeat_2x; 9.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 6.
KW EGF-like domain.
SQ SEQUENCE 2824 AA; 313502 MW; 27007746DD34E356 CRC64;

Query Match
Best Local Similarity 41.2%; Score 830; DB 13; Length 2824;
Matches 187; Conservative 71; Mismatches 113; Indels 140; Gaps 12;

QY 3 HAANGKRPASRAGR-----IPTSSPSLLPSAQLPSSHNPPVSC----- 44
Db 127 HAVLWGRSNTKSGRSSCLFSRANGLTDTDEHENTENGPLCHCSSASSSPDVPYPP 186
QY 45 -----QWPLDTSIQIMDTNPDSEFSPNLLRACSG---PQOASSGPPNHHQS 94

```

```

Db 187 SHAAQSQGRLLGNSGAQAQRDSEDEFGPNSFLVKTGSGNVCTPAAATANEFSQNH 246
QY 95 TLR-PPLPPPHNHTLSHHH-SSANSLNRNSLTNRSSQIHAPAPNDLA-----TTP 145
Db 247 RLRTPLPLFHSHPHQHTASIGLSRSNTYQSRN-----PSPAPTDSSAPNEGTSAD 302
QY 146 SVQLQDSWVLSNVPLETR-----HFLFKTSS 172
Db 303 SSSAQDNWLLNSNVPLETRNIAKQTFLTDQNFTEMDLATARRDGAYTDGHFLFK-PG 361
QY 173 GSTPLFSSSSPGYPLTSGTVTPPPRLPRNTFSKAPLKKPKSKYCWKAALSAIAA 232
Db 362 GTSPLYCTSPGYPLTSTSTVSPPPRPLPRNTFSRPAFSLKKPKYKCNWKAALSAI 421
QY 233 LLLAILLAYFI-----VPWSL----- 248
Db 422 VTLVFLAYFIAMHLFGLNHLQVQRQIYQITEDNTSGLHLPTDLGLPPLGNTGLEPPD 481
QY 249 -----KNSSIDSGAEVGRVTVQVPPGVFWRSQIHSIQPKFNISLGR 294
Db 482 RGRDDGKLDGFFPEDSFIDMGEIDVGRKVAQLIPPFIWRSQVFDHPVHLKFN 541
QY 295 DALFGVYIRRGLPSPSHQAQYDFMERLDGKREKSV-----VESPRRSIQTLVONEAVF 347
Db 542 DALVGIYGRGLPSPHTQDFVELLDGRLLSQGLPGLDPPFPAQORSLVPTSHDTGC 601
QY 348 VOYLDVGLWHLAFYNDGDKKEMVSFNTVVL 378
Db 602 IQYMDSGIWHLAYNDGKETEQVSLFTTAID 632

RESULT 9
QW7S7
ID QW7S7 PRELIMINARY; PRT; 2771 AA.
AC QW7S7;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NAR-2003 (TREMBlrel. 23, Last annotation update)
DE Ten-m4.
GN NRGL OR ODZ4 OR TEN-M4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c; TISSUE=Brain;
RA Ohashi T., Zhou X., Feng K., Richter B., Moergelin M., Perez M.T.,
RT "Mouse Ten-m/Odz is a new family of dimeric type II transmembrane
RT proteins expressed in many tissues.";
RL J. Cell Biol. 0:0-0(1999).
DR EMBL; AB025413; BAA77399.1; -.
DR HSSP; P16109; IFSB.
DR MGD; MGI:96083; Nrgl.
DR MGD; MGI:2447063; Odz4.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001258; NHL.
DR InterPro; IPR006530; YD.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF01436; NHL; 6.
DR SMART; SM00181; EGF; 6.
DR TIGRfams; TIGR01643; YD_repeat_2x; 7.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 7.
KW EGF-like domain.
SQ SEQUENCE 2771 AA; 308497 MW; 1492EE1EA0DBF0C CRC64;

Query Match
Best Local Similarity 39.3%; Score 793.5; DB 11; Length 2771;
Matches 173; Conservative 46; Mismatches 86; Indels 115; Gaps 9;

```

```

QY 69 NSYLLRACSGPQOASGSSGPPNHHQSOTLRPLPP-PHNHTLSHHH-SSANSNRNSLTNR 126
DB 149 NSNLTLDTEHENTETDHPSSSLQNHPLRTPPPPLPHAHTPNQHHASINSLARGNETPR 208
QY 127 RSQIHAPAPAPND-----LATTESVOLQDSWVNSNVPLETR----- 164
DB 209 SN-----PSPAPTDHSLSGEPAGAGAEPTHAQDNWLLNSNIPLETRNLGKQPFGLTLDQN 264
QY 165 -----HFEKTSSTGSPLEFSSSGPYPLTSGTYVTPPPRLPNTF 205
DB 265 LIEMDILSASRDGAYSDGHFLFK-PGTSPLCTTSPGYPLTSSVYSPPPRLPRSTF 323
QY 206 SRKAFLLKPKSKYCKWKAALSIAAALLAILLAYFI-----VPSWL- 248
DB 324 SRPAFLNLKPKSKYCKWKAALSIAISATVILLAYFVAMHLFGLNHLQPMQOMYE 383
QY 249 -----KNSSIDSGEAEVGR 263
DB 384 ITEDTASSWPVPTDVSILPSGGTGLTETPDRKGKGAEGKPSLFPEDSFIDSGEIDVGR 443
QY 264 VTQVPPGVFWSQIHSQPFKFNISLGDALFGYIIRGLPPSHQAQYDFMERLDGK- 322
DB 444 ASQKIPPGTFRWSQVFDHPVHLKFNYSLGKALVGIYGRKGLPPSHQTFDFVELLDGR 503
QY 323 -----EKWSVSPRERRSIOTLVQNEAVFYQYLDVGLWHLAFYNDGKDKEMVSFNTVLD 378
DB 504 LTQEARSLGPGQORSGVPPSPSHSETGTGFIQYLDGSIWHLAFYNDGKESEVVSFLATAIE 563

RESULT 10
P98202 PRELIMINARY; PRT; 560 AA.
AC P98202; 014667;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Downstream of CHOP protein (DOC-4) (Teneurin) (Fragment).
GN OD24 OR DOC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. OF GAMMA-HERGULIN FUSION PROTEIN.
RC TISSUE-BREAST CANCER;
RX MEDLINE-97472144; PubMed-9333014;
RA Schaefer G., Fitzpatrick V.D., Sliwkowski M.X.;
RT "Gamma-hergulin: a novel heregulin isoform that is an autocrine
growth factor for the human breast cancer cell line, MDA-MB-175.";
RL Oncogene 15:1385-1394 (1997).
RN [2]
RP CHROMOSOMAL TRANSLOCATION.
RX MEDLINE-99455251; PubMed-10523851;
RA Wang X.-Z., Jolicoeur E.M., Conte N., Chaffanet M., Zhang Y.,
RA Mozziconacci M.-J., Feiner H., Birnbaum D., Pebusque M.-J., Ron D.;
RT "Gamma-hergulin is the product of a chromosomal translocation fusing
the DOC4 and HGL/NRG1 genes in the MDA-MB-175 breast cancer cell
line.";
RL Oncogene 18:5718-5721(1999).
RN [3]
RP CHROMOSOMAL TRANSLOCATION.
RX MEDLINE-20065180; PubMed-10597312;
RA Liu X., Baker E., Eyre H.J., Sutherland G.R., Zhou M.;
RT "Gamma-hergulin: a fusion gene of DOC-4 and neuregulin-1 derived from
a chromosome translocation.";
RL Oncogene 18:7110-7114(1999).
CC 1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC 1- INDUCTION: BY STRESS.
CC 1- DISEASE: INVOLVED IN A RARE T(8;11) CHROMOSOMAL TRANSLOCATION THAT
FUZZES THE 5' END OF OD24 TO NRG1 (ISOFORM BETA3). THE PRODUCT OF
THIS TRANSLOCATION HAS FIRST BE THOUGHT TO BE AN ALTERNATIVE
SPLICED ISOFORM, CALLED GAMMA-HERGULIN. NOT DETECTED IN BREAST
CARCINOMA SAMPLES, INCLUDING DUCTAL, LOBULAR, MEDULLARY, AND

```

```

CC MUCINOUS HISTOLOGICAL TYPES, NEITHER IN OTHER BREAST CANCER CELL
CC LINES.
CC -1- SIMILARITY: TO D. MELANOGASTER TENN/ODZ.
DR EMBL: AF009227; AAC51756.1; ALT TERM.
KW Transmembrane; Chromosomal translocation.
FT DOMAIN 179 182 POLY-PRO.
FT TRANSMEM 346 366 POTENTIAL.
FT NON_TER 560 560
SQ SEQUENCE 560 AA; 61494 MW; 3927565E76C26C1D CRC64;

Query Match 39.2%; Score 790.5; DB 4; Length 560;
Best Local Similarity 41.8%; Pred. No. 1.1e-62;
Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;

QY 69 NSYLLRACSGPQOASGSSGPPNHHQSOTLRPLPP-PHNHTLSHHH-SSANSNRNSLTNR 126
DB 149 NSNLTLDTEHENTETDHPGQLNHARLTPPPPLSHAHTPNQHHASINSLARGNETPR 208
QY 127 RSQIHAPAPAPND--LATTP-----ESVOLQDSWVNSNVPLETR----- 164
DB 209 SN-----PSPAPTDHSLSGEPAGAGAEPTHAQDNWLLNSNIPLETRNLGKQPFGLTLDQN 264
QY 165 -----HFEKTSSTGSPLEFSSSGPYPLTSGTYVTPPPRLPNTF 205
DB 265 LIEMDILSASRDGAYSDGHFLFK-PGTSPLCTTSPGYPLTSSVYSPPPRLPRSTF 323
QY 206 SRKAFLLKPKSKYCKWKAALSIAAALLAILLAYFI-----VPSWL- 248
DB 324 ARPAFLNLKPKSKYCKWKAALSIAISATVILLAYFVAMHLFGLNHLQPMQOMYEIT 383
QY 249 -----KNSSIDSGEAEVGRRT 265
DB 384 EDTASSWPVPTDVSILPSGGTGLTETPDRKGKGTGKPSFFPEDSFIDSGEIDVGRAS 443
QY 266 QEVPPGVFWSQIHSQPFKFNISLGDALFGYIIRGLPPSHQAQYDFMERLDGK--- 322
DB 444 QKIPPGTFRWSQVFDHPVHLKFNYSLGKALVGIYGRKGLPPSHQTFDFVELLDGRLL 503
QY 323 --EKWSVSPRERRSIOTLVQNEAVFYQYLDVGLWHLAFYNDGKDKEMVSFNTVVL 377
DB 504 TQEARSLGTPGQORSGVPPSPSHSETGTGFIQYLDGSIWHLAFYNDGKESEVVSFLTAI 560

RESULT 11
Q9NZJ2 PRELIMINARY; PRT; 442 AA.
AC Q9NZJ2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE OD23 (Fragment).
GN OD23.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP Ben-Zur T., Feige E., Motro B., Wides R.;
RT "The mammalian Odz gene family: Homologs of a Drosophila pair rule
gene with expression implying distinct yet overlapping developmental
roles.";
RL Dev. Biol. 117:107-120(2000).
DR EMBL: AF195420; AAF28318.1; -.
FT NON_TER 442 442
SQ SEQUENCE 442 AA; 49073 MW; F4DD0DA48E66B75 CRC64;

Query Match 38.2%; Score 770; DB 4; Length 442;
Best Local Similarity 50.8%; Pred. No. 5.5e-61;
Matches 156; Conservative 36; Mismatches 55; Indels 60; Gaps 5;

QY 51 SNTSHQIMDTPDEEFSNLSYLLRACSGPQOASGSSGPPNHHQSOTLRPLPPPHNHTLSH 110

```

```

Db 151 SNSALTLTDEHENK-----SDSENEQASNOGOSTLQ-PLPPSHKQHSQAQ 195
QY 111 HSSANSLNRNSLTNRSSQIHA-PAPAPNDLATTPESVOLQDSWVLSNVPLETRHLEFK 169
Db 196 HPSITSLNRNSLTNRNSQSPAPPAALPAELQTTPEVOLQDSWVLSNVPLESRHLEFK 255
QY 170 TSSGSPLESSSSPGVPLTSGVYVTPPPRLLPRNTESRKAFKLUKPSKCSWKCAALSAL 229
Db 256 TGTGTPPLESTATPGYTMASSGVSYPTRPLPRNLTLSRAFAFKKSKSCWKCTALCAV 315
QY 230 AAALLAILLAYFI-----VPSWL----- 248
Db 316 GYSVLAAILLSFIAHHLGLNQLQOQTENDTFENGKVSNDTMTVTVSLPSGDNKGLGG 375
QY 249 ----KNSSDSGAEAYGRRVTQVPPGVFWRSQIHISQPFKFNISLGDALFGVYIRRG 305
Db 376 FTQENNTIDSGELDIGRAIQEIPPGIFWRSOLFIDQPOFLKFNISLQKDALIGVYGRKK 435
QY 306 LPPSHAQ 312
Db 436 LPPSHTQ 442

RESULT 12
Q9W6V6 PRELIMINARY; PRT; 2705 AA.
AC Q9W6V6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Teneurin-1.
GN TEN-1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99276585; PubMed=10341219;
RA Minet A.D., Rubin B.P., Tucker R.P., Baumgartner S.,
RA Chiquet-Ehrismann R.;
RT "Teneurin-1, a vertebrate homologue of the Drosophila pair-rule gene
RT ten-m, is a neuronal protein with a novel type of heparin-binding
RT domain.";
RL J. Cell Sci. 112:2019-2032(1999).
DR EMBL; AJ238613; CAB43098.1; -
DR HSSP; P00750; ITPG.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001258; NHL.
DR InterPro; IPR006530; YD.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF01436; NHL; 5.
DR SMART; SM00181; EGF; 6.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 7.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 7.
KW EGF-like domain.
SQ SEQUENCE 2705 AA; 302385 MW; 230F03D1999037D2 CRC64;

Query Match 37.0%; Score 745.5; DB 13; Length 2705;
Best Local Similarity 39.1%; Pred. No. 1e-57;
Matches 174; Conservative 56; Mismatches 120; Indels 95; Gaps 15;

QY 2 LHAANKRPSAEAGRIPTSPSLPSAQL-----PSSHNPPVSCOMPLLDNNTS 54
Db 91 LHGVSEHSYP-LEVGSVDVTETEGGASPDHALRMWRGMKSEHS-----LSSRAN 141
QY 55 HOLMDNPDEEFPNSVLLRACSGPQQAASSGPPNHHOSOTLR---PPLPPPHNHTLSHH 111
Db 142 SALSLSLTDHE-----RKSDG-----ENDMPGSPHNQFTFRPLPPPPPHACTCTRK 189

```

```

QY 112 -HSSANSLNRNSLTNRSSQIHAPAP-APNDLATTPESVOLQDSWVLSNVPLETRHLEFK 169
Db 190 PPAADSLQRRSMTT-RSQ---PSPAAPTPTQDSVHLHNSWVLSNIPLETRHLEFK 245
QY 170 TSSGSPLESSSSPGVPLTSGVYVTPPPRLLPRNTESRKAFKLUKPSKCSWKCAALSAL 229
Db 246 HGSSSAIFSAASQNYPLTSNTVYSPRPLRNPTESRPAFTFSKPYRCNWKCTALSAT 305
QY 230 AAALLAILLAYFI-----VPSWL----- 248
Db 306 ATVTVLLALLAYVIAVHLEGLTWQLQPVGOLYENGSKNGKAESTDDTTSYGKGVSD 365
QY 249 ----KNSSDSGAEAYGRRVTQVPPGVFWRSQIHISQPFKFNISLGDALFGV 300
Db 366 KTEKKVFQGRAIDTGEVEIGAQMOTIPPGFLFWRFQITHHVPVLYLKNISLAKDSILGI 425
QY 301 YTRGGLPPSHAQYDFMERLDGKREKSVSPRRRSIQTLVN-----EAVFYQYLDV 353
Db 426 YGRRNIPPTHTQDFVKLMDGKQ--LIKQEPKNSEEPQAAPRNLILTSLOETGFIEMDQ 483
QY 354 GLWHLAFYNDGKDKEMVSENTVVLD 378
Db 484 GAHMAFYNDGKKVEQVFLTTAIE 508

RESULT 13
Q8C8D2 PRELIMINARY; PRT; 536 AA.
AC Q8C8D2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Odd Oz/ten-m homolog 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK047423; BAC33054.1; -
SQ SEQUENCE 536 AA; 60081 MW; C9814627AE9B363F CRC64;

Query Match 36.6%; Score 739; DB 11; Length 536;
Best Local Similarity 41.4%; Pred. No. 4.5e-58;
Matches 170; Conservative 50; Mismatches 111; Indels 80; Gaps 11;

QY 33 LPSSHNPPVSCOMPLLDNNTSHQIMDTPDEE-----FSPNSYLLRACSGPQAQ 82
Db 128 MKSEHS-----SC-----LSSRANSALSLTDTHERKSDGNGFKFSPVCCDMEAPDASQD 179
QY 83 SSSGPPNHHQSOTLRPLPPPHNHTLSHH-HSSANSLNRNSLTNRSSQIHAPAPNDLA 141
Db 180 MOSSPHNQFTFRPLPPPPPHACTCARPPPVDSLQRRSMTT-RSQSPAPAPP--T 236
QY 142 TPESVOLQDSWVLSNVPLETRHFLTKTSSGSTPLFSSSSPGYPLTSGVYVTPPPRLLP 201
Db 237 STQDSVHLHNSWVLSNIPLETRHFLFKHSGSSAIFSAASQNYPLTSNTVYSPRPLP 296
QY 202 RNTFSRKAFLKPKSKYCSWKCAALSALIAALLAILLAYFIY-----PMSL----- 248
Db 297 RSTFSRPAFTENKPYRCNWKCTALSATAITVTALLAYVIVHLEGLTWQLQPVGQIYA 356
QY 249 -----KNSSDSGAEAYGRRVTQVPPGVFWR 274
Db 357 NGISNGNPCTESMDTTSYPIGGRVSDKSEKVKFGKRAIDTGEVDIGAQVMTIPPGLEW 416

```



```

QY 275 RSQIHISQOPLKFNISLGDALFGVYIRRGPPSHQAQYDFMERLDGKE-----KWS 326
DB 417 RFQITIHPIYKFNISLAKDSLGIYGRNIPHTPTQDFVKLMGDKQLVKQDSKSSD 476
QY 327 VVSPRRRSIQTLVQNEAVFQYLDVGLWHLAFYNDGKCKEMVSNFTVVL 377
DB 477 IQHSP--RNILTSLQ-ETGFIEYMQGPWYLAFYNDGKMEQVFLTTAI 524

RESULT 14
Q9UKZ4 PRELIMINARY; PRT: 2725 AA.
AC Q9UKZ4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Tenascin-M1.
GN TNM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20025749; PubMed=10556288;
RA Brandau O., Schuster V., Weiss M., Hellebrand H., Fink F.M.,
RA Kreczy A., Friedrich W., Strahm B., Niemeyer C., Belohradsky B.H.,
RA Meindl A.;
RT "Epstein-Barr virus-negative boys with non-Hodgkin lymphoma are
RT mutated in the SH2B1A gene, as are patients with x-linked
RT lymphoproliferative disease (XLP).";
RL Hum. Mol. Genet. 8:2407-2413(1999).
DR EMBL; AF100772; AAF04723.1;
DR HSSP; P00750; ITPG.
DR GENE; HGNC:8117; ODZ1.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001258; NHL.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR006530; YD.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF01436; NHL; 6.
DR SMART; SM00181; EGF; 6.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 7.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
KW EGF-like domain.
SQ SEQUENCE 2725 AA; 304892 MW; 6F383B64C1B2E1B1 CRC64;

Query Match 36.6%; Score 739; DB 4; Length 2725;
Best Local Similarity 41.3%; Pred. No. 3 9e-57;
Matches 170; Conservative 51; Mismatches 113; Indels 78; Gaps 10;

QY 33 LPSSHNPPVSCOMPLDLSNTSHQIMDTPDEE-----FSPNSYLLRACSGPQA 82
DB 128 MKSEHS-----SC-----LSSRANSALSLTDTHERKSDGNGFKFSPVCCDMEAGAGSTQD 179
QY 83 SSSGPPNHHSQSLTRPLPPPHNHTLSHH--HSSANSLNRLNRLNRQIHAPAPNDLA 141
DB 180 VQSSPHNQFTFRPLPPPPPPHACTCARPPPAADSLQRKSMTT-RSQSPAAPAPP--T 236
QY 142 TTPEVOLQSWLNSVNLPLETRHFLFKTSSTGTPFLFSSSGPGVPLTSGTVYTPPRLLP 201
DB 237 STQDSVHLNSVWLNINPIETRHSLFKHSGSSAIFSAASQNYPLTSNTVYSPPPRLP 296
QY 202 RNTFSRAFKLKPSKYSCKAALSIAAALLAILLAYFI-----VPWSL----- 248
DB 297 RSTFSRPAFTFNKPYRCNKKCTALSATAITVTALLAVYIAVHLFGLTWQLQPVBEGL 356
QY 249 -----KNSSIDSGAEVGRVTVQVPPGV 272
DB 357 ANGISNGNPGETESMDTTYSPIGGRVSDKSEKKVQKGRADTGEVDIGAQMOTIPPLG 416

```

```

DB 357 YANGVSKNGRGTESMDTTYSPIGKVSDEKSEKKVQKGRADTGEVDIGAQMOTIPPLG 416
QY 273 FWSQIHISQOPLKFNISLGDALFGVYIRRGPPSHQAQYDFMERLDGKE-----KWS 326
DB 417 RFQITIHPIYKFNISLAKDSLGIYGRNIPHTPTQDFVKLMGDKQLVKQDSKGS 476
QY 327 VVSPRRRSIQTLVQNEAVFQYLDVGLWHLAFYNDGKCKEMVSNFTVVL 378
DB 477 DDTQHSRNLILTSQ-ETGFIEYMQGPWYLAFYNDGKMEQVFLTTAI 527

RESULT 15
Q9WTS4 PRELIMINARY; PRT: 2731 AA.
AC Q9WTS4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Ten-m1.
GN ODZ1 OR TEN-M1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Balb/c;
RA Ohashi T., Zhou X., Feng K., Richter B., Moergelin M., Perez M.T.,
RA Su W., Chiquet-Ehrismann R., Rauch U., Faessler R.;
RT "Mouse Ten-m/od2 is a new family of dimeric type II transmembrane
RT proteins expressed in many tissues.";
RL J. Cell Biol. 0:0-0(1999).
DR EMBL; AB025410; BAA77396.1;
DR HSSP; P00750; ITPG.
DR MGD; MGI:1345185; Odz1.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001258; NHL.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR006530; YD.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF01436; NHL; 5.
DR SMART; SM00181; EGF; 7.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 7.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
KW EGF-like domain.
SQ SEQUENCE 2731 AA; 305792 MW; 9129FA4CFE4A7770 CRC64;

Query Match 36.6%; Score 737.5; DB 11; Length 2731;
Best Local Similarity 40.9%; Pred. No. 5 3e-57;
Matches 169; Conservative 52; Mismatches 111; Indels 81; Gaps 11;

QY 33 LPSSHNPPVSCOMPLDLSNTSHQIMDTPDEE-----FSPNSYLLRACSGPQA 82
DB 128 MKSEHS-----SC-----LSSRANSALSLTDTHERKSDGNGFKFSPVCCDMEAPADSQD 179
QY 83 SSSGPPNHHSQSLTRPLPPPHNHTLSHH--HSSANSLNRLNRLNRQIHAPAPNDLA 141
DB 180 MQSSPHNQFTFRPLPPPPPPHACTCARPPPTVDSLQRKSMTT-RSQSPAAPAPP--T 236
QY 142 TTPEVOLQSWLNSVNLPLETRHFLFKTSSTGTPFLFSSSGPGVPLTSGTVYTPPRLLP 201
DB 237 STQDSVHLNSVWLNINPIETRHFLFKHSGSSAIFSAASQNYPLTSNTVYSPPPRLP 296
QY 202 RNTFSRAFKLKPSKYSCKAALSIAAALLAILLAYFI-----VPWSL----- 248
DB 297 RSTFSRPAFTFNKPYRCNKKCTALSATAITVTALLAVYIAVHLFGLTWQLQPVGIY 356
QY 249 -----KNSSIDSGAEVGRVTVQVPPGV 273
DB 357 ANGISNGNPGETESMDTTYSPIGGRVSDKSEKKVQKGRADTGEVDIGAQMOTIPPLG 416

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 14:29:58 ; Search time 21 Seconds
(without alignments)
767.640 Million cell updates/sec

Title: US-09-584-411c-22
Perfect score: 2017
Sequence: 1 MLHANKGRPSAAGRP...NDGDKEMVSNFTVLDGTI 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 790.5 | 39.2 | 560 | 3 | US-08-891-845-4 |
| 2 | 790.5 | 39.2 | 560 | 4 | US-09-514-573-4 |
| 3 | 790.5 | 39.2 | 768 | 3 | US-08-891-845-2 |
| 4 | 790.5 | 39.2 | 768 | 4 | US-09-514-573-2 |
| 5 | 658 | 32.6 | 501 | 3 | US-08-891-845-10 |
| 6 | 658 | 32.6 | 501 | 4 | US-09-514-573-10 |
| 7 | 122 | 6.0 | 872 | 1 | US-08-491-357-3 |
| 8 | 122 | 6.0 | 872 | 3 | US-08-968-633-3 |
| 9 | 122 | 6.0 | 872 | 3 | US-09-190-466-3 |
| 10 | 122 | 6.0 | 872 | 5 | PCT-US96-10823-3 |
| 11 | 117 | 5.8 | 612 | 4 | US-08-894-454-163 |
| 12 | 117 | 5.8 | 657 | 1 | US-08-264-534-34 |
| 13 | 117 | 5.8 | 657 | 1 | US-08-083-590A-13 |
| 14 | 117 | 5.8 | 657 | 1 | US-08-465-500-34 |
| 15 | 117 | 5.8 | 657 | 2 | US-08-346-148-34 |
| 16 | 117 | 5.8 | 657 | 3 | US-08-532-384-13 |
| 17 | 117 | 5.8 | 657 | 3 | US-08-893-828-34 |
| 18 | 117 | 5.8 | 2471 | 1 | US-08-185-432-16 |
| 19 | 117 | 5.8 | 2471 | 1 | US-08-083-590A-19 |
| 20 | 117 | 5.8 | 2471 | 3 | US-08-532-384-19 |
| 21 | 117 | 5.8 | 2471 | 4 | US-08-899-232-1 |
| 22 | 116.5 | 5.8 | 681 | 1 | US-08-083-590A-18 |
| 23 | 116.5 | 5.8 | 681 | 2 | US-08-346-128-37 |
| 24 | 116.5 | 5.8 | 681 | 3 | US-08-532-384-18 |
| 25 | 116.5 | 5.8 | 1078 | 1 | US-08-264-534-32 |
| 26 | 116.5 | 5.8 | 1078 | 1 | US-08-083-590A-11 |
| 27 | 116.5 | 5.8 | 1078 | 1 | US-08-465-500-32 |

| | | | | | | |
|----|-------|-----|------|---|-------------------|-------------------|
| 28 | 116.5 | 5.8 | 1078 | 2 | US-08-346-128-32 | Sequence 32, Appl |
| 29 | 116.5 | 5.8 | 1078 | 3 | US-08-532-384-11 | Sequence 11, Appl |
| 30 | 116.5 | 5.8 | 1078 | 3 | US-08-893-828-32 | Sequence 32, Appl |
| 31 | 116.5 | 5.8 | 2556 | 1 | US-08-083-590A-20 | Sequence 20, Appl |
| 32 | 116.5 | 5.8 | 2556 | 3 | US-08-532-384-20 | Sequence 20, Appl |
| 33 | 116 | 5.8 | 2972 | 4 | US-09-579-181-2 | Sequence 2, Appl |
| 34 | 116 | 5.8 | 3118 | 4 | US-09-579-181-1 | Sequence 1, Appl |
| 35 | 113.5 | 5.6 | 1481 | 2 | US-08-616-844-40 | Sequence 40, Appl |
| 36 | 113.5 | 5.6 | 1481 | 2 | US-08-599-654-40 | Sequence 40, Appl |
| 37 | 113.5 | 5.6 | 1481 | 3 | US-08-944-868A-40 | Sequence 40, Appl |
| 38 | 113.5 | 5.6 | 1481 | 3 | US-08-944-423A-40 | Sequence 40, Appl |
| 39 | 113.5 | 5.6 | 1481 | 3 | US-08-944-486-40 | Sequence 40, Appl |
| 40 | 113 | 5.6 | 1706 | 2 | US-08-459-568-2 | Sequence 2, Appl |
| 41 | 113 | 5.6 | 1706 | 2 | US-08-399-411-2 | Sequence 2, Appl |
| 42 | 113 | 5.6 | 1706 | 3 | US-08-516-859A-2 | Sequence 2, Appl |
| 43 | 113 | 5.6 | 1706 | 4 | US-09-586-472-2 | Sequence 2, Appl |
| 44 | 113 | 5.6 | 1706 | 4 | US-09-528-706-2 | Sequence 2, Appl |
| 45 | 112 | 5.6 | 472 | 3 | US-08-348-518C-5 | Sequence 5, Appl |

ALIGNMENTS

RESULT 1
US-08-891-845-4
; Sequence 4, Application US/08891845
; Patent No. 6096873
; GENERAL INFORMATION:
; APPLICANT: Schaefer, Gabriele M.
; APPLICANT: Sliwowski, Mark
; TITLE OF INVENTION: Gamma-Herregulin
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,845
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/021640
; FILING DATE: 07/12/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P1043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 560 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-891-845-4

Query Match 39.2%; Score 790.5; DB 3; Length 560;
Best Local Similarity 41.5%; Pred. No. 5e-64;
Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;
Qy 69 NSYLLRACSGPQQSSGPPNHHQSOSTLRPLPP-PHNHTLSHHH;SSANSLNRLNRLNR 126
DB 149 NSNLTLDTEHENTETDHPGGLQNHARLTTPPPPLSHARTPNQHHAAASINSILNRGNFTPR 208

Query Match 39.2%; Score 790.5; DB 3; Length 768;
Best Local Similarity 41.5%; Pred. NO. 7.8e-64;
Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;
US-09-514-573-2

QY 69 NSYLRLACSGPQOASSGPPNHHQSOSTLRPPLPP-PHNHTLSHHH-SSANSLNRLNRLNR 126
DB 149 NSNLTLDTEHENTETDHPGGLQNHARLRTPPPLSHAHTPNQHHASINSLNRCNFTPR 208
QY 127 RSQIHAPAPAND--LATTP-----ESVQLQDSWVLSNSVPLETR----- 164
DB 209 SN-----PSPAPTHSLSGEPAGGAQEPAAHAQENWLLNSNIPLETRNLGKQPFGLTLODN 264
QY 165 -----HFLFTYSGSTPLFSSSPGYPLTSGVTYTPPRLPRNTF 205
DB 265 LIEMDILGASRDGAYS DGHFLEK-PGCTSPFLCTTSPGYPLTSTVYSPPPRLPRSTF 323
QY 206 SRKAFKLKPKSKYCWKCAALSAIAAALLAILLAYFI-----VPNSL----- 248
DB 324 ARPAENLKKPKSKYCNWKAALSAIVISATLVILLAYFVAMHLFGLNHLQPMWEGOMYEIT 383
QY 249 -----KNSIDSGEAEVGRRTV 265
DB 384 EDTASSWPVPTDVSILYPSGGTGLETPDRKGKGTTEGKPSFFPEDSFIDSGEIDVGRAS 443
QY 266 QEVPPGVFWRSQIHSQPFKFNISLQKDALFGVYIRRGPLPPSHAQYDFMERLDGK--- 322
DB 444 QKIPPGTFWRSQVFDHVPVHLKFNVLGKAALVGIYGRKGLPPSHTQDFVELLDGRLL 503
QY 323 --EKWSVVEPRRERSIOTLVQNEAVFVQYLDVGLWHLAFYNDGKDKEMVSFNTVVL 377
DB 504 TQEARSLGTPRQSRGTVPPSSSHETGFIQYLDGSIWHLAFYNDGKESEVVSFLTTAI 560

RESULT 4

US-09-514-573-2
; Sequence 2, Application US/09514573
; Patent No. 6500941
; GENERAL INFORMATION:
; APPLICANT: Schaefer, Gabriele M.
; APPLICANT: Sliwowski, Mark
; TITLE OF INVENTION: Gamma-Heregulin
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,573
; FILING DATE: 28 FEB 2000
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/891845
; FILING DATE: 10 JULY 1997
; APPLICATION NUMBER: 60/021640
; FILING DATE: 07/12/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P1043
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 768 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-09-514-573-2

Query Match 39.2%; Score 790.5; DB 4; Length 768;
Best Local Similarity 41.5%; Pred. NO. 7.8e-64;
Matches 173; Conservative 48; Mismatches 83; Indels 113; Gaps 10;
US-09-514-573-2

QY 69 NSYLRLACSGPQOASSGPPNHHQSOSTLRPPLPP-PHNHTLSHHH-SSANSLNRLNRLNR 126
DB 149 NSNLTLDTEHENTETDHPGGLQNHARLRTPPPLSHAHTPNQHHASINSLNRCNFTPR 208
QY 127 RSQIHAPAPAND--LATTP-----ESVQLQDSWVLSNSVPLETR----- 164
DB 209 SN-----PSPAPTHSLSGEPAGGAQEPAAHAQENWLLNSNIPLETRNLGKQPFGLTLODN 264
QY 165 -----HFLFTYSGSTPLFSSSPGYPLTSGVTYTPPRLPRNTF 205
DB 265 LIEMDILGASRDGAYS DGHFLEK-PGCTSPFLCTTSPGYPLTSTVYSPPPRLPRSTF 323
QY 206 SRKAFKLKPKSKYCWKCAALSAIAAALLAILLAYFI-----VPNSL----- 248
DB 324 ARPAENLKKPKSKYCNWKAALSAIVISATLVILLAYFVAMHLFGLNHLQPMWEGOMYEIT 383
QY 249 -----KNSIDSGEAEVGRRTV 265
DB 384 EDTASSWPVPTDVSILYPSGGTGLETPDRKGKGTTEGKPSFFPEDSFIDSGEIDVGRAS 443
QY 266 QEVPPGVFWRSQIHSQPFKFNISLQKDALFGVYIRRGPLPPSHAQYDFMERLDGK--- 322
DB 444 QKIPPGTFWRSQVFDHVPVHLKFNVLGKAALVGIYGRKGLPPSHTQDFVELLDGRLL 503
QY 323 --EKWSVVEPRRERSIOTLVQNEAVFVQYLDVGLWHLAFYNDGKDKEMVSFNTVVL 377
DB 504 TQEARSLGTPRQSRGTVPPSSSHETGFIQYLDGSIWHLAFYNDGKESEVVSFLTTAI 560

RESULT 5

US-08-891-845-10
; Sequence 10, Application US/08891845
; Patent No. 6096873
; GENERAL INFORMATION:
; APPLICANT: Schaefer, Gabriele M.
; APPLICANT: Sliwowski, Mark
; TITLE OF INVENTION: Gamma-Heregulin
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,845
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/021640
; FILING DATE: 07/12/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P1043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994

```
;
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 501 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
;
; US-08-891-845-10
;
; Query Match 32.6%; Score 658; DB 3; Length 501;
; Best Local Similarity 47.5%; Pred. No. 5.8e-52;
; Matches 132; Conservative 33; Mismatches 47; Indels 66; Gaps 4;
;
; QY 165 HLEFKTSSGTPLEFSSSGYPLTSGTYVTPPPRLPRNTFSKAFKLLKPKSKYCSWKCA 224
; DB 17 HELFK-PGTSPLFCTSPGYPLTSGTYVTPPPRLPRNTFSKAFKLLKPKSKYCSWKCA 75
;
; QY 225 ALSAIAAALLALLAYFI-----VPWSL----- 248
; DB 76 ALSAIVISATLVILLAYFVAMHLFGLNHLQPMEGOMYEITDASSWPVPTDVSLYPSG 135
;
; QY 249 -----KNSSIDSGAEVGRVTVQVPPGVFWRSQIHISQPO 284
; DB 136 GTGLETPDRKGKGTGKPPSSFPEDSFIDSGEIDVGRRASQKIPPGTFWRQVFDHPV 195
;
; QY 285 FLKFNISLGDALFGYVIRGLPPSHQAQYDFMERLDGK-----EKSVVSPRERSIQT 339
; DB 196 HLKFNYSGLKAALVGIYGRKGLPPSHQTFDFVLLDGRLLTQEARSLGTPRQSRGTVP 255
;
; QY 340 LVONEAVFYQYLDVGLWHLAFYNDGKEMVSEFNTVVL 377
; DB 256 PSSHETGFIQYLDGSIWHLAFYNDGKESEVVSFLTAI 293
;
; RESULT 6
; US-09-514-573-10
; Sequence 10, Application US/09514573
; Patent No. 6500941
; GENERAL INFORMATION:
; APPLICANT: Schaefer, Gabriele M.
; APPLICANT: Sliwowski, Mark
; TITLE OF INVENTION: Gamma-Herregulin
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,573
; FILING DATE: 28 FEB 2000
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/891845
; FILING DATE: 10 JULY 1997
; APPLICATION NUMBER: 60/021640
; FILING DATE: 07/12/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P1043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 10:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 501 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
;
; US-09-514-573-10
;
; Query Match 32.6%; Score 658; DB 4; Length 501;
; Best Local Similarity 47.5%; Pred. No. 5.8e-52;
; Matches 132; Conservative 33; Mismatches 47; Indels 66; Gaps 4;
;
; QY 165 HLEFKTSSGTPLEFSSSGYPLTSGTYVTPPPRLPRNTFSKAFKLLKPKSKYCSWKCA 224
; DB 17 HELFK-PGTSPLFCTSPGYPLTSGTYVTPPPRLPRNTFSKAFKLLKPKSKYCSWKCA 75
;
; QY 225 ALSAIAAALLALLAYFI-----VPWSL----- 248
; DB 76 ALSAIVISATLVILLAYFVAMHLFGLNHLQPMEGOMYEITDASSWPVPTDVSLYPSG 135
;
; QY 249 -----KNSSIDSGAEVGRVTVQVPPGVFWRSQIHISQPO 284
; DB 136 GTGLETPDRKGKGTGKPPSSFPEDSFIDSGEIDVGRRASQKIPPGTFWRQVFDHPV 195
;
; QY 285 FLKFNISLGDALFGYVIRGLPPSHQAQYDFMERLDGK-----EKSVVSPRERSIQT 339
; DB 196 HLKFNYSGLKAALVGIYGRKGLPPSHQTFDFVLLDGRLLTQEARSLGTPRQSRGTVP 255
;
; QY 340 LVONEAVFYQYLDVGLWHLAFYNDGKEMVSEFNTVVL 377
; DB 256 PSSHETGFIQYLDGSIWHLAFYNDGKESEVVSFLTAI 293
;
; RESULT 7
; US-08-491-357-3
; Sequence 3, Application US/08491357
; Patent No. 5716782
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, JoAnne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/491,357
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
```

```
; ANTI-SENSE: NO
US-08-491-357-3

Query Match
Best Local Similarity 6.0%; Score 122; DB 1; Length 872;
Matches 79; Conservative 35; Mismatches 85; Indels 150; Gaps 23;

QY 9 RKPAEAGRPPTPPSSPSLLPSAQLPSS-HNP-PPVSCOMPLLDSTNTHQIMDTPDEEF 66
DB 68 KKPAARG--PGPPATPPO--POPSPQGVHTFVPPASQYSPMLP--TAYQ---PQPD--- 115
QY 67 SPNSYLLRACSGPQOASSGPPNHHSSQSTLPPPLPPPHNHTLSH-----HHSSANSLNRN 121
DB 116 --NYLVPTPSKTQOGLYQAFGNQFQS-----PPAKQTSYKQTPHHS----- 158
QY 122 SLTNRROIHAPAPNDLATTPESVQLQDSWVLSNVPLETRHFLFKTSSGSTPLFSS 181
DB 159 -----PPSPA-TDLYQVP-----P 171
QY 182 SPGYPL-----TSGT---VYTPPPRLPRNTFSRKAFKLPKPKSKYCSWKCAALSAIA 230
DB 172 GPGSPAQDIYQVPPSAGTGHDYQVPPSL-----DTRSWEGTKPP----- 211
QY 231 AALLLAILLAYFIVPWSLKNSSI-DSGEA-----VGRVTOEVPP--GVF-- 273
DB 212 -----AKVVVTRVCGQYVVEASQAQDEYDTPRHLLAPGSQDIYDVPVRGLLPN 262
QY 274 -WRSQIHISQOFLKFNISLGDALFGY-----IRRLGPPS--HAQYD 314
DB 263 QYGOEYDTPDMAVKG--PNGRDPLLDYDVPVPSVEKGLPPSNHHSYVD 309

RESULT 8
US-08-968-633-3
; Sequence 3, Application US/08968633
; Patent No. 6100384
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, JoAnne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,633
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid

; STRANDEDNESS: NO. 6100384 Relevant
; TOPOLOGY: NO. 6100384 Relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-968-633-3

Query Match
Best Local Similarity 6.0%; Score 122; DB 3; Length 872;
Matches 79; Conservative 35; Mismatches 85; Indels 150; Gaps 23;

QY 9 RKPAEAGRPPTPPSSPSLLPSAQLPSS-HNP-PPVSCOMPLLDSTNTHQIMDTPDEEF 66
DB 68 KKPAARG--PGPPATPPO--POPSPQGVHTFVPPASQYSPMLP--TAYQ---PQPD--- 115
QY 67 SPNSYLLRACSGPQOASSGPPNHHSSQSTLPPPLPPPHNHTLSH-----HHSSANSLNRN 121
DB 116 --NYLVPTPSKTQOGLYQAFGNQFQS-----PPAKQTSYKQTPHHS----- 158
QY 122 SLTNRROIHAPAPNDLATTPESVQLQDSWVLSNVPLETRHFLFKTSSGSTPLFSS 181
DB 159 -----PPSPA-TDLYQVP-----P 171
QY 182 SPGYPL-----TSGT---VYTPPPRLPRNTFSRKAFKLPKPKSKYCSWKCAALSAIA 230
DB 172 GPGSPAQDIYQVPPSAGTGHDYQVPPSL-----DTRSWEGTKPP----- 211
QY 231 AALLLAILLAYFIVPWSLKNSSI-DSGEA-----VGRVTOEVPP--GVF-- 273
DB 212 -----AKVVVTRVCGQYVVEASQAQDEYDTPRHLLAPGSQDIYDVPVRGLLPN 262
QY 274 -WRSQIHISQOFLKFNISLGDALFGY-----IRRLGPPS--HAQYD 314
DB 263 QYGOEYDTPDMAVKG--PNGRDPLLDYDVPVPSVEKGLPPSNHHSYVD 309

RESULT 9
US-09-196-466-3
; Sequence 3, Application US/09196466
; Patent No. 6124434
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, JoAnne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,466
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/491,357
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
```

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-196-466-3

Query Match 6.0%; Score 122; DB 3; Length 872;
Best Local Similarity 22.6%; Pred. No. 0.015;
Matches 79; Conservative 35; Mismatches 85; Indels 150; Gaps 23;
QY 9 RKPSAAGRIPIPTSPSLLPSAQLPSS--HNP--PPVSCQMPLLDSNTSHQIMDTNDEEF 66
DB 68 KKAAPG--PGPATPPQ--PQSLPGQVHTPPVPASQYSPMLP--TAYQ---PQPD--- 115
QY 67 SPNSYLLRACSGPQQAASSGPPNHHQSQTLLRPPPHNHTLSH-----HHSSANSLSNRN 121
DB 116 --NVILVPTPSKTQOGLYQAPGNPQFS-----PPAKQTSTFSKQTPHHS----- 158
QY 122 SLTNRRSQIHAPAPANDLATTPEVSQLODSWVLSNVPLETRHFLFKTSSGSTPLFSSS 181
DB 159 -----FPSPA--TDLYQVP-----P 171
QY 182 SPGYPL-----TSGT---VYTPPRLPRNTFSRKAFKLLKPKSKYCSWKCAALSAIA 230
DB 172 GFGSPAQDIYQVPPSAGTGHDIYQVPPSL-----DTRSWEGTKPP----- 211
QY 231 AALLAILLAYFTVPMWSLKNSSI--DSGEAE-----VGRVTOEVP--GVF-- 273
DB 212 -----AKVVPTRVGGYVYEASQAQDEYDTPRHLLAPGSQDIYDVPVRCGLLPN 262
QY 274 --WRSQIHISQPFLENISLGDALFGVY-----IRGLPPS--HAQYD 314
DB 263 QYQGEYDTPPMAVGK--PNGRDPDLLDYDVPVPSVEKGLPPSNHHSYVD 309

RESULT 10

PCT-US96-10823-3
Sequence 3, Application PC/TUS9610823
GENERAL INFORMATION:
APPLICANT: Golemis, Erica A.
APPLICANT: Law, Susan
APPLICANT: Estojak, JoAnne
TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
TITLE OF INVENTION: ALTERATIONS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10823
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US96-10823-3

Query Match 6.0%; Score 122; DB 5; Length 872;
Best Local Similarity 22.6%; Pred. No. 0.015;
Matches 79; Conservative 35; Mismatches 85; Indels 150; Gaps 23;
QY 9 RKPSAAGRIPIPTSPSLLPSAQLPSS--HNP--PPVSCQMPLLDSNTSHQIMDTNDEEF 66
DB 68 KKAAPG--PGPATPPQ--PQSLPGQVHTPPVPASQYSPMLP--TAYQ---PQPD--- 115
QY 67 SPNSYLLRACSGPQQAASSGPPNHHQSQTLLRPPPHNHTLSH-----HHSSANSLSNRN 121
DB 116 --NVILVPTPSKTQOGLYQAPGNPQFS-----PPAKQTSTFSKQTPHHS----- 158
QY 122 SLTNRRSQIHAPAPANDLATTPEVSQLODSWVLSNVPLETRHFLFKTSSGSTPLFSSS 181
DB 159 -----FPSPA--TDLYQVP-----P 171
QY 182 SPGYPL-----TSGT---VYTPPRLPRNTFSRKAFKLLKPKSKYCSWKCAALSAIA 230
DB 172 GFGSPAQDIYQVPPSAGTGHDIYQVPPSL-----DTRSWEGTKPP----- 211
QY 231 AALLAILLAYFTVPMWSLKNSSI--DSGEAE-----VGRVTOEVP--GVF-- 273
DB 212 -----AKVVPTRVGGYVYEASQAQDEYDTPRHLLAPGSQDIYDVPVRCGLLPN 262
QY 274 --WRSQIHISQPFLENISLGDALFGVY-----IRGLPPS--HAQYD 314
DB 263 QYQGEYDTPPMAVGK--PNGRDPDLLDYDVPVPSVEKGLPPSNHHSYVD 309

RESULT 11

US-08-894-454-163
Sequence 163, Application US/08894454
Patent No. 6544784
GENERAL INFORMATION:
APPLICANT: VAN DEN VEN, W.J.M.
APPLICANT: SCHOENMAKERS, H.F.P.M.
TITLE OF INVENTION: MULTIPLE-TUMOR ABERRANT GROWTH
TITLE OF INVENTION: GENES
NUMBER OF SEQUENCES: 164
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Webb Law Firm
STREET: 700 Koppers Building, 436 Seventh Avenue
CITY: Pittsburgh
STATE: PA
COUNTRY: USA
ZIP: 15219-1818
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,454
FILING DATE: 15-AUG-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP/00716
FILING DATE: 19-FEB-1996
APPLICATION NUMBER: 95200390.3
FILING DATE: 17-FEB-1995
APPLICATION NUMBER: 95201951.1
FILING DATE: 14-JUL-1995

ATTORNEY/AGENT INFORMATION:
NAME: Johnson, Barbara E
REGISTRATION NUMBER: 31,198
REFERENCE/DOCKET NUMBER: 702-971100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 412-471-8815
TELEFAX: 412-471-4094
TELEX:
INFORMATION FOR SEQ ID NO: 163:
SEQUENCE CHARACTERISTICS:
LENGTH: 612 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-894-454-163

RESULT 12
US-86-264-534-34
; Sequence 34, Application US/08264534
; Patent No. 5648464
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Sgyr
; TITLE OF INVENTION: Human No. 56484
; TITLE OF INVENTION: In Toporythmic
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americ
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,534
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 07/695,189
; FILING DATE: 03-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-000

```

;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: 212 780-3090
;
; TELEFAX: 212 8698864/9741
;
; TELEX: 66141 PENNIE
;
; INFORMATION FOR SEQ ID NO: 34:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 657 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: unknown
;
; MOLECULE TYPE: protein
;
; US-08-264-534-34

```

RESULT 13
US-08-083-590A-13
; Sequence 13, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083.590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090

TELEFAX: 212 869864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 657 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-083-590A-13

Query Match 5.8%; Score 117; DB 1; Length 657;
Best Local Similarity 22.1%; Pred. No. 0.029;
Matches 79; Conservative 50; Mismatches 111; Indels 118; Gaps 21;
QY 7 KGRKPSAEAGRPPTSPSLPSAQ-LPSSHNPVPVSCQMPLLDNTSHQIMDTPDEE 65
DB 288 KSRPSSAKS---TMPSTLNLAKEDAKGSRKKSLSEKVLSESVTLSPVDSLE--- 341
QY 66 FSPNSY-----LLRACGPOQASSGPPNHHQSSTL-----RPPLP--- 101
DB 342 -SPHTVSDTSSPMITSPGILQASPNMLATAAPAPVHAQHALSFNLHEMQLAHGA 400
QY 102 ----PPHNTLSHH-----SSANSLNRSLNRRSQIHAPAPAPNDLATTTPESVOLQD 151
DB 401 STVLPSVQSLSHHHIVSPGSGSAGLSR-----LH-PVPVPAD----- 438
QY 152 SWVLNSNVPLETRH-----FLFKTSSGSTPLFSSSPGYPLTSGTYTP-----PP---R 198
DB 439 -WMNRMEVN-ETQYNEMFGWLAPABGTHPGIAPOS--RPPEKHHITTPREPLPIVTFQ 494
QY 199 LLPRNFTFSKAPKLPKPSKYCSWKCAALSAIAAALLAILLAYFIVPWSLKNSSIDSGEA 258
DB 495 LIPKGSIAQPA-GAPQOSTCP-----PAVAGP-----LPTMYQIP----- 529
QY 259 EYGRRTVQVPPGVFWRSQIHISQ---POFLKFNISLGDALFGVYIRRGLPSPSHAQY 313
DB 530 EMARLPSVAFPTAMPPOQDQVAQTILPAYHPPASVGK-----YTPPSQHSY 578

RESULT 14
US-08-465-500-34
Sequence 34, Application US/08465500
Patent No. 5789195
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Muskavitch, Marc A.T.
APPLICANT: Fehon, Richard G.
APPLICANT: Rebay, Ilaria
APPLICANT: Blauweller, Cristine M.
APPLICANT: Shepard, Scott B.
TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,500
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-034

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 657 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-465-500-34
Query Match 5.8%; Score 117; DB 1; Length 657;
Best Local Similarity 22.1%; Pred. No. 0.029;
Matches 79; Conservative 50; Mismatches 111; Indels 118; Gaps 21;
QY 7 KGRKPSAEAGRPPTSPSLPSAQ-LPSSHNPVPVSCQMPLLDNTSHQIMDTPDEE 65
DB 288 KSRPSSAKS---TMPSTLNLAKEDAKGSRKKSLSEKVLSESVTLSPVDSLE--- 341
QY 66 FSPNSY-----LLRACGPOQASSGPPNHHQSSTL-----RPPLP--- 101
DB 342 -SPHTVSDTSSPMITSPGILQASPNMLATAAPAPVHAQHALSFNLHEMQLAHGA 400
QY 102 ----PPHNTLSHH-----SSANSLNRSLNRRSQIHAPAPAPNDLATTTPESVOLQD 151
DB 401 STVLPSVQSLSHHHIVSPGSGSAGLSR-----LH-PVPVPAD----- 438
QY 152 SWVLNSNVPLETRH-----FLFKTSSGSTPLFSSSPGYPLTSGTYTP-----PP---R 198
DB 439 -WMNRMEVN-ETQYNEMFGWLAPABGTHPGIAPOS--RPPEKHHITTPREPLPIVTFQ 494
QY 199 LLPRNFTFSKAPKLPKPSKYCSWKCAALSAIAAALLAILLAYFIVPWSLKNSSIDSGEA 258
DB 495 LIPKGSIAQPA-GAPQOSTCP-----PAVAGP-----LPTMYQIP----- 529
QY 259 EYGRRTVQVPPGVFWRSQIHISQ---POFLKFNISLGDALFGVYIRRGLPSPSHAQY 313
DB 530 EMARLPSVAFPTAMPPOQDQVAQTILPAYHPPASVGK-----YTPPSQHSY 578

RESULT 15
US-08-346-128-34
Sequence 34, Application US/08346128
Patent No. 5856441
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon et al.
TITLE OF INVENTION: Human No. 5856441ch And Delta, Binding Domains
TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/346,128
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/879,038
FILING DATE: 30-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-009

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 657 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-346-128-34

Query Match 5.8%; Score 117; DB 2; Length 657;
Best Local Similarity 22.1%; Pred. No. 0.029;
Matches 79; Conservative 50; Mismatches 111; Indels 118; Gaps 21;

| | | | |
|----|-----|---|-----|
| QY | 7 | KGRKPSAEAGRPITPTSSPLPSAQ-LPSSHPPVSCOMPLLDSTSHQIMDTNPDEE | 65 |
| Db | 288 | KSRPSPAKS---TMPTSLNLAKKADAKGSRKKSLSEKVLSESVTLSPVDSLE--- | 341 |
| QY | 66 | FSPNSY-----LLRACSGPQQAASSGPPNHHQSSTL-----RPPLP--- | 101 |
| Db | 342 | -SPHTVSDTSSPMITSPGILQASPNPLATAAPAPVHAQHLSFSLNHEMQLAHGA | 400 |
| QY | 102 | -----PPHNTLSHHH-----SSANSLNRNSLTNRRSQIHAPAPAPNDLATTPESVQLQD | 151 |
| Db | 401 | STVLPVSQQLSHHHIVSPGSGSAGLSR-----LH-PVPVPAD----- | 438 |
| QY | 152 | SMVLNSNVPLETRH-----FLKTSSTGTPLFSSSPGYPLTSGTVYTP-----PP---- | 198 |
| Db | 439 | -WNRMEVN-ETQYNEMFCWVLAPAEHTHPGIAPOS--RPPECKHITTPREPLPIVTFQ | 494 |
| QY | 199 | LIPRNTFSKAFKLPKSKYCSWKCAALSAIAAALLAILLAYFIVPWSLKNSSIDSGEA | 258 |
| Db | 495 | LIPKGSIAQPA-GAPQPSTCP-----PAVAGP-----LPTMYQIP----- | 529 |
| QY | 259 | EVGRRYVQEPGVPFVRSQIHISQ----POFLKFNISLGDALFGVYIRGLPPSHAQY | 313 |
| Db | 530 | EMARLPSVAFPTAMMPQDQVAQTILPAYHPFPASVGR-----YPTPPSQHSY | 578 |

Search completed: August 26, 2003, 14:32:51

Job time : 23 secs

THIS PAGE BLANK (user)